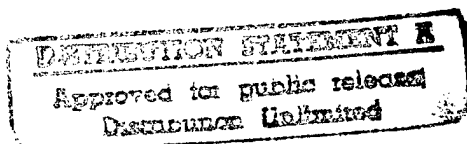


N7714745 1083



QUALITY INSPECTED

19970814 064

A Service of:



National Aeronautics and
Space Administration

Scientific and Technical
Information Program Office
Center for AeroSpace Information

NOTICE

**THIS REPRODUCTION WAS MADE FROM THE BEST AVAILABLE
COPY OF WHICH A NUMBER OF PAGES WERE OF POOR
REPRODUCTION QUALITY.**

(NASA-TM-X-73170) TWELFTH ANNUAL CONFERENCE
ON MANUAL CONTROL (NASA) 1001 P HC A99/MF
A01 CSCL 05E

N77-14745

Unclass

G3/54 58298

**NASA TECHNICAL
MEMORANDUM**

NASA TM X-73,170

NASA TM X-73,170

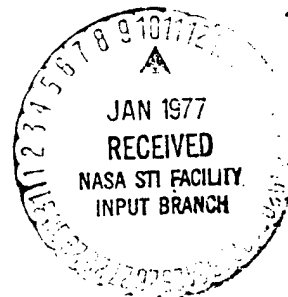
TWELFTH ANNUAL CONFERENCE ON MANUAL CONTROL

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

and

Ames Research Center
Moffett Field, California 94035

May 1976



DTIC QUALITY INSPECTED 3

1. Report No. NASA TM X-73,170	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle TWELFTH ANNUAL CONFERENCE ON MANUAL CONTROL		5. Report Date	
		6. Performing Organization Code	
7. Author(s)		8. Performing Organization Report No. A-6762	
9. Performing Organization Name and Address Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801, and Ames Research Center, NASA, Moffett Field, CA 94035		10. Work Unit No. 505-09-33	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D. C. 20546		13. Type of Report and Period Covered Technical Memorandum	
		14. Sponsoring Agency Code	
15. Supplementary Notes Proceedings of a meeting held at Coordinated Science Laboratory, University of Illinois, Urbana, Illinois, May 25-27, 1976.			
16. Abstract This volume contains a compilation of written versions of papers presented at the Twelfth Annual Conference on Manual Control in a meeting held at the Coordinated Science Laboratory, University of Illinois, May 25-27, 1976. Eleven main topics were discussed during this three-day conference. These were covered in sessions on Multi-task Decision Making, Attention Allocation and Workload Measurement, Displays and Controls, Nonvisual Displays, Tracking and other Psychomotor Tasks, Automobile Driving, Handling Qualities and Pilot Ratings, Remote Manipulation, System Identification, Control Models, and Motion and Visual Cues. Sixty-five papers are included with presentations on results of analytical studies to develop and evaluate human operator models for a range of control task, vehicle dynamics and display situations; results of tests of physiological control systems and applications to medical problems; and on results of simulator and flight tests to determine display, control and dynamics effects on operator performance and workload for aircraft, automobile, and remote control systems.			
17. Key Words (Suggested by Author(s)) Manual control, Physiological control, Man-machine systems, Displays, Vehicle handling qualities, Operator performance, Flight management systems, Simulation		18. Distribution Statement Unlimited STAR Category - 54	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 1020	22. Price* \$25.25

FOREWORD

This volume contains the proceedings of the Twelfth Annual Conference on Manual Control held at the Coordinated Science Laboratory of the University of Illinois at Urbana-Champaign from May 25 to 27, 1976. This report contains complete manuscripts of most of the papers presented at the meeting.

This was the twelfth in a series of conferences dating back to December 1964. These earlier meetings and their proceedings are listed below:

First Annual NASA-University Conference on Manual Control, The University of Michigan, December 1964. (Proceedings not printed.)

Second Annual NASA-University Conference on Manual Control, MIT, February 28 to March 2, 1966, NASA SP-128.

Third Annual NASA-University Conference on Manual Control, University of Southern California, March 1-3, 1967, NASA SP-144.

Fourth Annual NASA-University Conference on Manual Control, The University of Michigan, March 21-23, 1968, NASA SP-192.

Fifth Annual NASA-University Conference on Manual Control, MIT, March 27-29, 1969, NASA SP-215.

Sixth Annual Conference on Manual Control, Wright-Patterson AFB, April 7-9, 1970.

Seventh Annual Conference on Manual Control, University of Southern California, June 2-4, 1971, NASA SP-281.

Eighth Annual Conference on Manual Control, University of Michigan, Ann Arbor, Michigan, May 17-19, 1972.

Ninth Annual Conference on Manual Control, Massachusetts Institute of Technology, May 23-25, 1973.

Tenth Annual Conference on Manual Control, Wright-Patterson AFB, April 9-11, 1974.

Eleventh Annual Conference on Manual Control, NASA-Ames Research Center, May 21-23, 1975, NASA, TM X-62,464.

PRECEDING PAGE BLANK NOT FILMED

CONFERENCE CHAIRMAN

William B. Rouse
University of Illinois

PROGRAM COMMITTEE

Renwick E. Curry
Massachusetts Institute of Technology

Rudolph G. Mortimer
University of Illinois

Richard W. Pew
Bolt Beranek and Newman

Thomas B. Sheridan
Massachusetts Institute of Technology

Robert C. Williges
University of Illinois

PUBLICATION COMMITTEE

Melvin Sadoff
NASA - Ames Research Center

Thomas E. Wempe
NASA - Ames Research Center

CONTENTS

Session I. MULTI-TASK DECISION MAKING

Chairman: R. W. Allen

	Page
1. Review of the Symposium on Monitoring Behavior and Supervisory Control (Berchtesgaden, F. R. Germany, March 8-12, 1976) by T. B. Sheridan	3
2. Human Control and Monitoring - Models and Experiments, by P. H. Wewerinke	14
3. An Experimental Situation for Study of Pilot Interaction With Automated Airborne Decision Making Systems, by W. B. Rouse, Y-Y. Chu, and R. S. Walden	39
4. An Experimental Situation for Study of Human Decision Making in Multi-Process Monitoring, by W. B. Rouse and J. S. Greenstein . .	45
5. The Psychophysics of Random Processes, by R. E. Curry and T. Govindaraj	50

Session II. ATTENTION ALLOCATION AND WORKLOAD MEASUREMENT

Chairman: R. G. Mortimer

6. A Novel Approach to the Cross-Adaptive Auxiliary Task, by A. R. Ephrath	63
7. Is There An Optimum Workload in Manual Control, by W. L. Verplank .	72
8. Task Interference in Multi-Axis Aircraft Stabilization, by E. D. Onstott	80
9. Results of a Display Design Technique Applied to VTOL Aircraft, by W. C. Hoffman, R. E. Curry, D. L. Kleinman (This paper was presented but not received for inclusion in the proceedings.)	
10. Telling a Computer How a Human Has Allocated His Attention Between Control and Monitoring Tasks, by K. D. Enstrom and W. B. Rouse . .	104
11. The Use of Event-Related Potentials in the Enhancement of System Performance, by C. D. Wickens, J. Isreal, G. McCarty, D. Gopher, and E. Donchin	124
12. Pupillometric Measurement of Cognitive Workload, by J. Beatty . . .	135

Session III. DISPLAYS AND CONTROLS

Chairman: R. W. Allen

13. Left-Brain/Right-Brain and Symbolic/Analogic Human Operator
Output Compatibility, *by W. L. Verplank* 147
14. A Regression Approach to Generate Aircraft Predictor
Information, *by P. D. Gullaher, R. A. Hunt, and R. C. Williges.* . . 156
15. Introduction to a Coordinated Cockpit Display Concept,
by D. L. Baty 166
16. Implications of a Mixture of Aircraft With and Without Traffic
Situation Displays for Air Traffic Management, *by J. Kreifeldt,
L. Parkin, P. Rothschild, and T. Wempe.* 179
17. Investigating the Use of a Moving Map Display and a Horizontal
Situation Indicator in Simulated Powered-Lift Short-Haul
Operations, *by W. F. Clement.* 201
18. Air-to-Ground Visual Display System, *by D. L. Kugel and
R. V. Gressang.* 225
19. In-Flight Simulation Study of Decoupled Longitudinal Controls for
the Approach and Landing of a STOL Airplane, *by E. Seckel and
B. Feinreich.* 244
20. High Acceleration Cockpit Displays and Controls,
by J. A. Townsend and N. J. Kropewnicki 271

Session IV. NONVISUAL DISPLAYS

Chairman: W. H. Lewison

21. Design of an English Language Data Base Question-Answering
System, *by D. L. Waltz.* 283
22. Effects of Linguistic Redundancy on Pilot's Comprehension of
Synthesized Speech, *by C. A. Simpson.* 294
23. Effects of Linguistic Redundancy on Synthesized Cockpit Warning
Message Comprehension and Concurrent Time Estimation,
by S. G. Hart and C. A. Simpson 309
24. Tactual Commands for Pilot Flare Training, *by R. D. Gilson, and
R. W. Ventola* 322
25. Two-Dimensional Compensatory Tracking With Tactile Displays,
by H. P. Schmid, G. A. Bekey, and J. B. Reswick 332

26. A Simulator Study of Helicopter Pilot Workload Reduction Using a Tactile Display, by R. S. Dunn, R. D. Gilson, and P. Sun. . . . 355
27. Control Augmentation and Workload Reduction by Kinesthetic Information From the Manipulator, by S. J. Merhav and O. Ben Ya'acov. 361

Session V. TRACKING AND OTHER PSYCHOMOTOR TASKS

Chairman: D. L. Baty

28. Time Optimal Control of an Undamped Harmonic Oscillator: Evidence for Biases and Schemata, by R. J. Jagacinski, M. W. Burke, and D. P. Miller 383
29. Continuous Compensatory Audio Manual Tracking, by R. G. Costello. . 406
30. Eye Tracking: Horizontal vs. Vertical, by J. Hornseth, G. Stanley, J. Porterfield, and P. Carson 422
31. Head Tracking. A Fatigue Study, by J. Hornseth, G. Stanley, and P. Carson 430
32. A Comparison of In-Flight and Ground-Based Pitch Attitude Tracking Experiments, by M. F. C. van Cool and H. A. Mooij. . . . 443
33. Effects of Headgear and Visual Angle on Head Rotation Spectral Characteristics, by D. K. Shirachi, D. L. Monk, and J. H. Black, Jr. 455
34. Human Engineering Laboratory Helicopter Observer Performance Studies, by J. A. Barnes. 469
35. Interaction Between Warning Signal and Fore-Period in Simple Foot Reaction Time, by N. S. Nataraj and J. M. Howard 477

Session VI. AUTOMOBILE DRIVING

Chairman: G. A. Bekey

36. A Simulator for Driving Research, by R. G. Mortimer and R. L. Adkins. 493
37. Alcohol Effects on Performance of Driver Decision-Making Tasks in a Laboratory Simulation, by R. W. Allen, S. H. Schwartz, and D. T. McRuer
(This paper was presented but not received for inclusion in the proceedings.)

38. Expected Value Analysis of Driver Decision Making Under Alcohol,
by *S. H. Schwartz and R. W. Allen*
(This paper was presented but not received for inclusion in
the proceedings.)
39. Spectral Analysis of the Driver-Vehicle System in Car Following,
by *R. Herman, R. Rothery, and R. Rule*
(This paper was presented but not received for inclusion in
the proceedings.)

Session VII. HANDLING QUALITIES AND PILOT RATINGS

Chairman: *S. J. Merhav*

40. A Statistical Approach to the Analysis of the Pilot Behavior on
Multiloop Systems, by *N. Goto* 501
41. Application of a Model-Based Flight Director Design Technique
to a Helicopter Hover Task, by *R. A. Hess*
(This paper was presented but not received for inclusion in
the proceedings.)
42. A Method for Generating Numerical Pilot Opinion Ratings Using
the Optimal Pilot Model, by *R. A. Hess*
(This paper was presented but not received for inclusion in
the proceedings.)
43. Ride Quality Sensitivity to SAS Control Law and to Handling
Quality Variations, by *P. A. Roberts, D. K. Schmidt, and
R. L. Swaim* 528
44. A Unified Theory for Pilot Opinion Rating, by *R. H. Smith* 542
45. Longitudinal Flying Qualities in the Landing Approach,
by *W. W. Rickard*. 559
46. Worth Assessments of Approach to Landing, by *R. E. Curry*. 585

Session VIII. REMOTE MANIPULATION

Chairman: *T. B. Sheridan*

47. A Computer-Controlled Machining and Manipulating Device,
by *T. C. Woo and J. M. Paul* 595
48. Mechanical Design and Computer Configuration in the Computer-
Aided Manipulator Control Problem, by *P. Coiffet, J. Vertur,
and E. Dombre* 607
49. A Preliminary Evaluation of Manual Control Tasks Associated With
the Space Shuttle Remote Manipulator System, by *L. Reid* 624

50. Modeling a Manipulation Task of Variable Difficulty,
by J. W. Hill and S. J. Matthews. 639
51. A Comparison of Master-Slave and Rate Control in the Presence
of a Transmission Time Delay, by G. P. Starr. 661

Session IX. SYSTEM IDENTIFICATION

Chairman: L. Reid

52. A Comparison of Techniques for Identifying Human Operator
Dynamics Utilizing Time Series Analysis, by K. Tanaka,
N. Goto, and K. Washizu 673
53. System Identification via a Microcomputer, by R. Mekei
and A. Singh. 694
54. PID (Proportional Integral Derivative) Modeling Techniques Applied
to Studies of Motion and Peripheral Display Effects on Human
Operator Performance, by D. W. Fepperger and A. M. Junker 703
55. Sensitivity Analysis of Motion and Peripheral Display Effects on
Tracking Performance, by D. W. Fepperger and A. M. Junker 719

Session X. CONTROL MODELS

Chairman: S. Baron

56. Low Visibility Landing Simulation Experiments, by R. V. Gressang,
J. J. Pollard, and D. L. Kugel. 733
57. The Application of Pilot Modeling to the Study of Low Visibility
Landing, by W. W. Harrington. 748
58. Stability of the Pilot-Aircraft System in Maneuvering Flight,
by J. R. Broussard and R. F. Stengel. 778
59. Analysis of a Manned Anti-Aircraft Artillery (AAA) System Using a
Proportional-Integral-Derivative (P-I-D) Structure Optimal Control
Gunner Model, by A. V. Phatak and K. M. Kessler 795
60. Analysis of Controls and Displays for a Terminal Controlled
Vehicle (Abstract), by W. H. Levison and S. Baron 819
61. A Model of the EMG-Torque-Movement Relationships for Knee
Extension, by E. Dombre, G. A. Zekey, and J. Perry. 820
62. A Model of the Reflex Arc and Analysis of the Behavior of the
Peripheral Receptors During Clonus, by J. E. Dollman,
A. M. Iannone, and L. T. Andrews. 836

Session XI. MOTION AND VISUAL CUES

Chairman: R. C. Williges

63. A Pilot Model With Visual and Motion Cues, *by R. E. Curry, L. R. Young, W. C. Hoffman, and D. L. Kugel* 851
64. The Evaluation of Peripheral Display Units in a Fixed Base Simulator, *by L. Reid and A. J. Fraser*. 855
65. Roll Axis Tracking Improvement Resulting From Peripheral Vision Motion Cues, *by T. E. Moriarty, A. M. Junker, and D. R. Price* . . . 868
66. Comparative Effects of Motion and Peripheral Displays on Roll Axis Tracking, *by A. M. Junker*
(This paper was presented but not received for inclusion in the proceedings.)
67. Use of Motion Cues in Steady-State Tracking, *by W. H. Levison* . . . 895
68. The Influence of Loss of Visual Cues on Pilot Performance During the Final Approach and Landing Phase of a Remotely Piloted Vehicle Mission, *by J. C. Howard*. 918
69. Simulation of Conventional Aircraft Approach and Landing: The Effect of Withholding Motion or Instrument Cues, *by R. V. Gressang* 932
70. Effects of Differing Motion System Drives on Simulation of Approach and Landing, *by J. J. Pollard*. 941
71. Thresholds of Motion and Perception Measured in a Flight Simulator, *by R. J. A. W. Hosman and J. C. van der Vaart*. 956
72. Biomechanical Response and Manual Tracking Performance in Sinusoidal, Sum-of-Sines, and Random Vibration Environments, *by W. H. Levison*. 985

SESSION I
MULTI-TASK DECISION MAKING

Chairman: R. W. Allen

REVIEW OF THE SYMPOSIUM ON MONITORING BEHAVIOR AND SUPERVISORY CONTROL

(BERCHTESGADEN, F.R. GERMANY, MARCH 8-12, 1976)*

by Thomas B. Sheridan

Massachusetts Institute of Technology
Cambridge, Ma. 02139

SUMMARY

The motivation for this meeting was the fact that the role of the human operator in manned systems is changing from that of a continuous in-the-loop controller to that of a monitor and supervisor. The rapid development of the computer is the primary force which is causing this change, especially the fact of marked decrease in cost and size, along with associated new and more sophisticated developments in software and display technology.

To deal with these issues, over 100 participants from fifteen countries assembled during 8-12 March, 1976, at Berchtesgaden, F.R. Germany, in the Bavarian Alps. Thirty-five papers were presented in three categories: vehicle control (aircraft, automobiles and urban mass transit, ships); process control (industrial manufacture, nuclear reactors); and general models of monitoring and supervisory control.

Intensive workshop discussions involving all the participants were held throughout the symposium to define better the human operator's role changes, the concomitant changes that are necessary (new priorities) in research, the associated problems in implementation, and required new interdisciplinary and institutional arrangements.

INTRODUCTION: THE NATURE OF SUPERVISORY CONTROL

Figure 1 characterizes the trend in controlling complex systems. Figure 1a represents control of a modern aircraft, where at the lowest level of control and on a short time scale, thrust, pitch, roll and yaw are stabilized relative to gust disturbances and reference commands coming from guidance logic. At a next level, the vehicle's course is controlled relative to prevailing winds and reference commands from the navigation logic. Finally, the navigation process can be automated, in part, by ground-based as well as on-board computers. At each level in the multi-loop control hierarchy the

* Dr. Gunnar Johannsen of Forschungsinstitut für Anthropotechnik, Meckenheim, Germany, was Co-director along with the author. The meeting was sponsored by the Special Programmes Panel on Human Factors, Scientific Affairs Div., NATO. and by the Federal Republic of Germany. The papers and workshop reports will soon be available from Plenum Press in a book entitled Monitoring Behavior and Supervisory Control.

MONITOR & SUPERVISE:

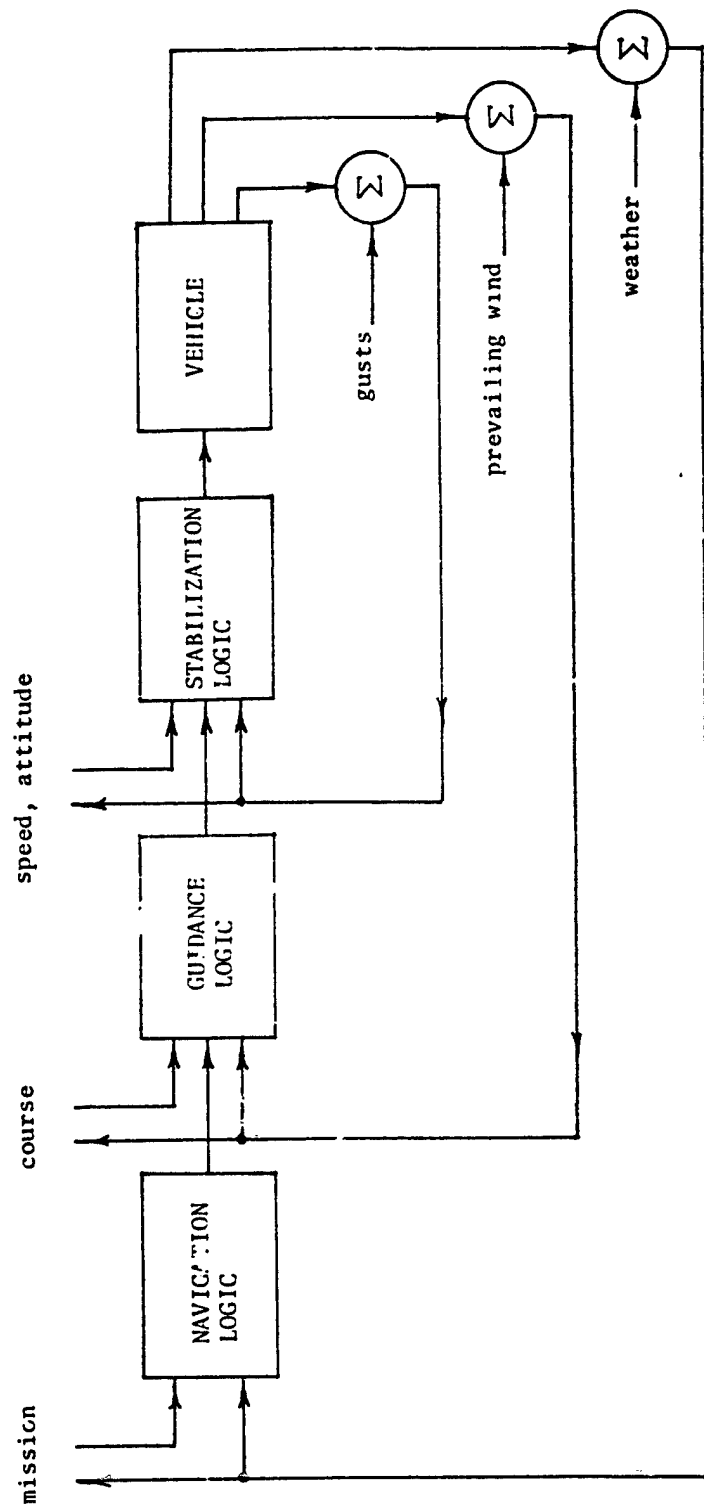


Figure 1a. Supervisory Control of Aircraft

human operator can make inputs. At the lowest level he can provide continuous stick-pedal commands to the inner loop (and does, of course, during critical take-off and landing maneuvers). At the middle level he can provide intermittent course changes by resetting course relative to a computer-navigation base. At the highest level he can modify the flight plan or mission in major ways, in conjunction with persons on the ground and/or other aircraft. At each higher level the nature of his control is different, responses becoming more intermittent (longer time between responses) and involving more monitoring and more interaction with computers and with other persons.

Actually it was probably in spacecraft where the multi-level control first was implemented to any degree of sophistication and where one might say we first saw "supervisory control". It is rather arbitrary to set the number of levels in the control hierarchy at three (except that the conjunction of the three words "navigation", "guidance" and "stabilization" has come into common use in the aerospace industry). In the Apollo spacecraft one can list many more levels of control for some functions, especially if one considers nested control loops in computer software, layers of backup or abort modes, etc.

Turning to a different kind of system, the chemical plant, one finds a similar hierarchy or nesting of control loops, with successively longer time constants, successively less continuous real-time automatic control, and successively more human interaction with higher level control. Figure 1b suggests three levels of control associated with: "valve control" to regulate instant-by-instant temperature and pressure; "mixing logic" to effect longer term mixing procedures; and "plant management" to control the major steps of producing the given product.

A third example (Figure 1c) is found with the industrial robot, as used for manipulating and assembling discrete parts on the production line. At the lowest level conventional servomechanisms control instantaneous forces and positions of the actuators for each degree of freedom. At a middle level a computer decides what sequences of positions and forces for the component degrees of freedom to command in order to achieve certain elementary manipulations ("therbligs") which are programmed from above, like: reach, grasp, insert, release, etc. The highest level the control system coordinates is the robot's accomplishment of various tasks in conjunction with the production line, special parts feeders, other robots or human workers, etc. Again, most of the human effort is in monitoring and supervising at the highest control level, though in case of emergencies or for maintenance or repair the human operator may intervene into automatic loops and apply direct control at lower levels.

Figure 2 illustrates supervisory control from the viewpoint of the human operator. He observes displays which may present pictorial, graphic, or alphanumeric information. In the near future he may also listen to computer-generated speech information. He operates hand controls which may be of the symbolic type (keyboard) or analogic type (joystick, light pen). In the near future he may speak in specialized code and be recognized by computer.

The characterization in Figure 2 shows a "local" control loop in which the human operator can test his plans and programs before committing them to action. This is analogous to the way the chief executive uses planning

MONITOR & SUPERVISE:

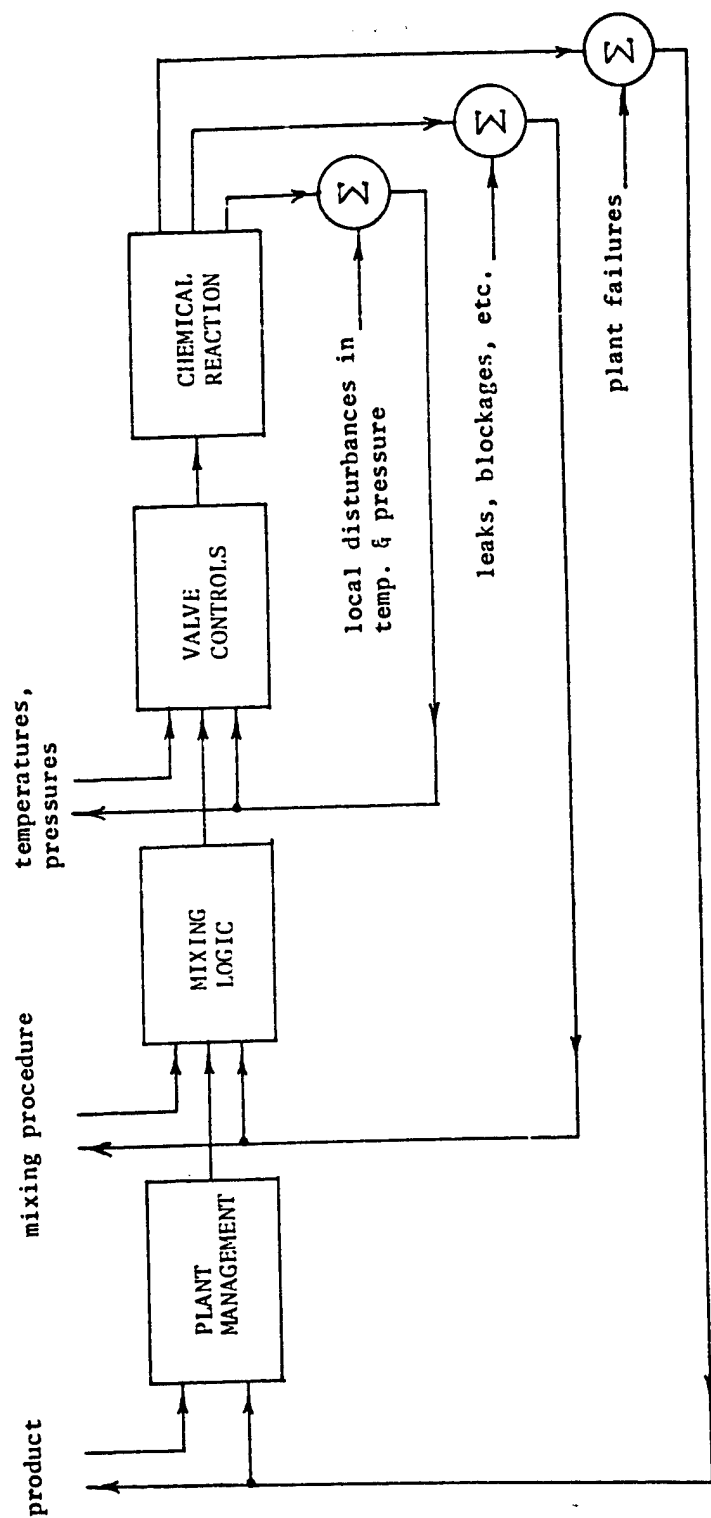


Figure 1b. Supervisory Control of
Chemical Plant

MONITOR & SUPERVISE:

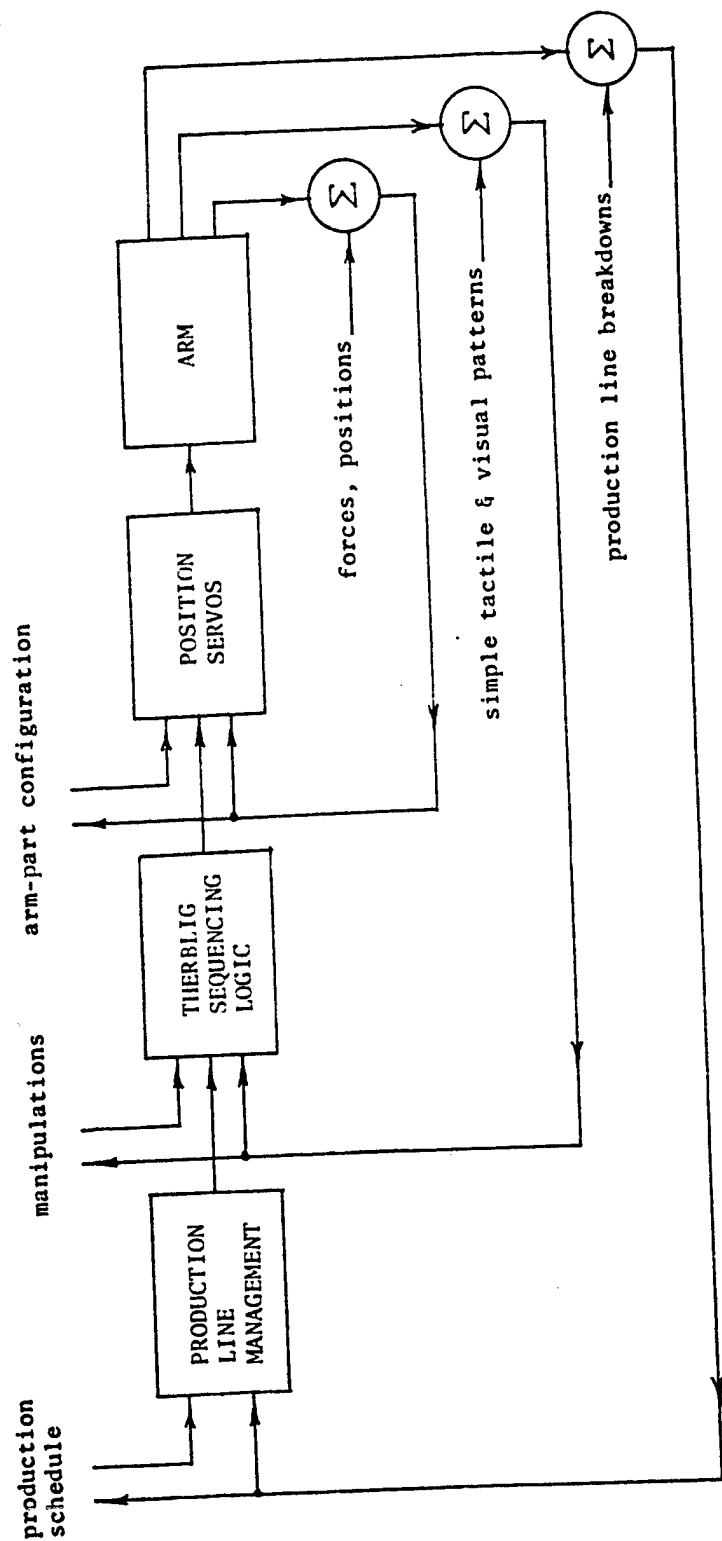


Figure 1c. Supervisory Control of Industrial Robot

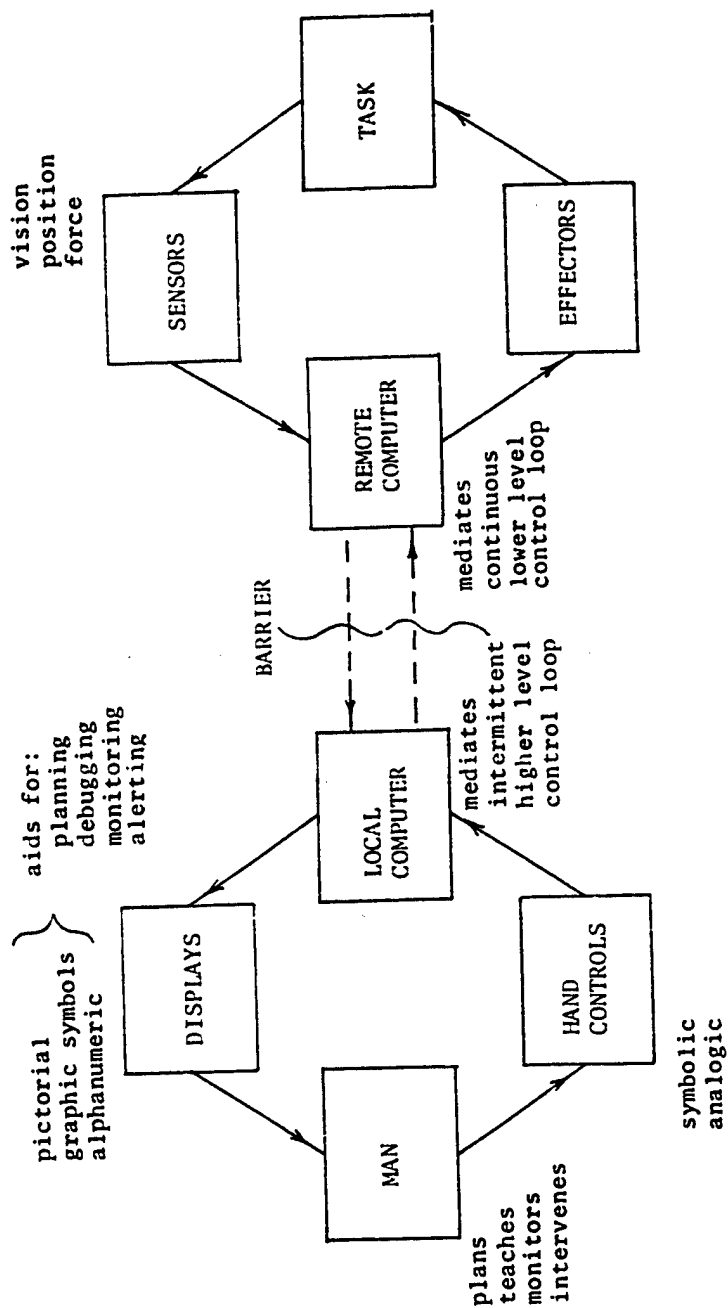


Figure 2. Paradigm of Supervisory Control

staff in the managerial hierarchy of any large organization. After both plans and communications (teaching or programming in this case) have been worked out, the chief executive orders his line organization into action (via the supervisory link in Figure 2). He then monitors what happens as his subordinates (in this case the "remote" control loop, with its own sensors, delegated decision-making powers, and capability to effect its own actions) take over, reassuming control in emergency or when his instructions have been executed.

EMPHASES AND POINTS OF CONSENSUS

The following, in the opinion of the author, constitute the major highlights, emphases of the papers presented at the Symposium, and points of consensus which emerged from the workshops. The reader is referred to the full papers and workshop reports, forthcoming from Plenum Press.

Multiplicity of Interfaces

Understanding monitoring behavior and supervisory control requires observation and analysis at a multiplicity of system interfaces. Figure 3 (after E. Edwards of the University of Loughborough, England) is a convenient way of portraying these interfaces: the conventional man-machine interface between a human operator ("liveware") and the hardware; a linguistic interface between him and the software; a physiological interface between him and the environment; and a social interface between him and other persons in the system. Edwards called this the "SHELL" model for obvious reasons.

Use of Simulators

The papers and workshops affirmed the increasing use of simulators in complex man and computer-controlled systems. A first reason is that, because of the rich electronic intercommunication of component subsystems in simulators and the availability of computers intrinsic to the system, it is relatively easy to do measurements, data storage, data correlation, and even on-line modeling of human behavior. A second reason for simulation is that it is the best way (perhaps the only viable way) to provide human operators exposure to low probability events - which is one of the major reasons that human operators are included in such systems.

Workload

There was a great deal of discussion about human operator workload. Four alternative definitions of workload were evident. A first definition is in terms of what the task demands are for "satisfactory performance" (required reduction in uncertainty, positioning speed or accuracy, forces required, etc.). A second is in terms of the level of effort or difficulty as perceived (subjectively) by the operator. A third is in terms of what the

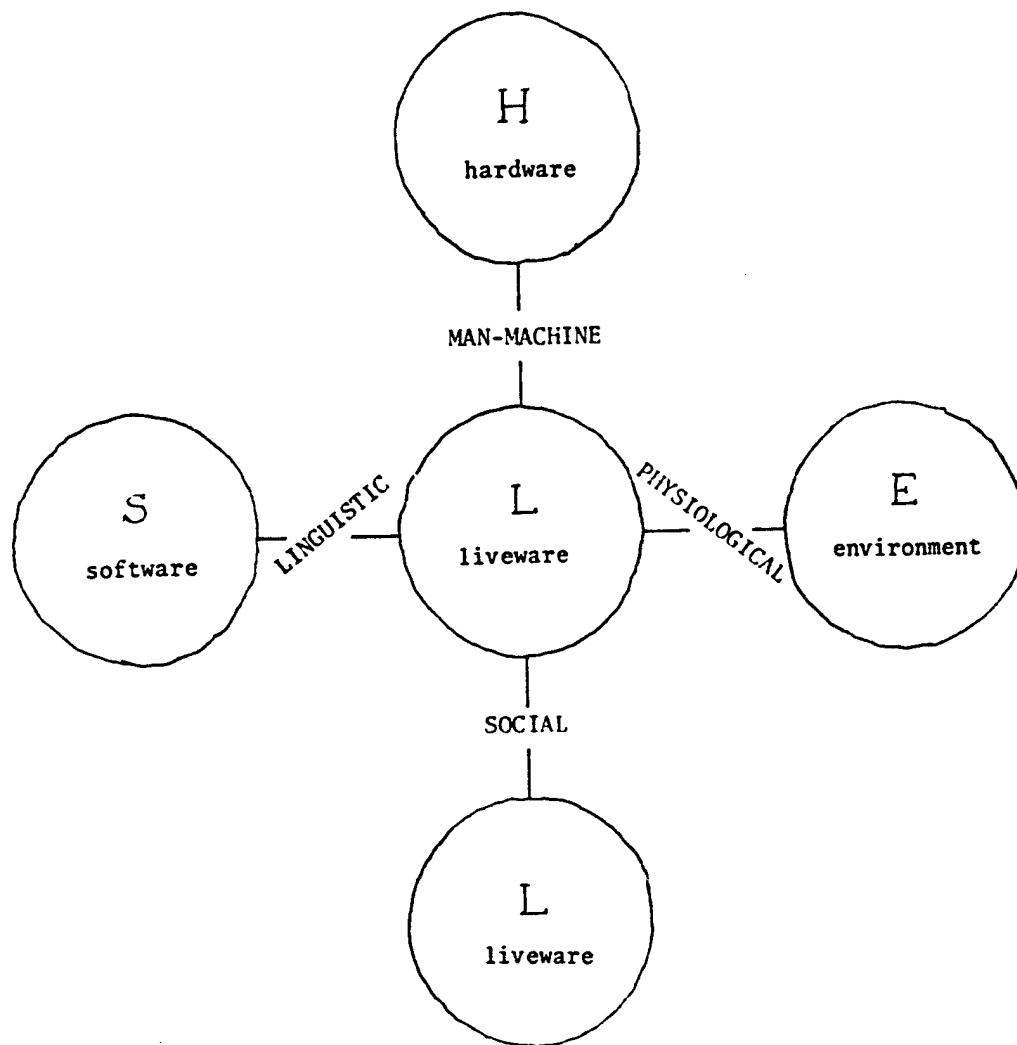


Figure 3. SHELL Model of Interfaces

operator actually accomplishes, or some objective measure of his decision success, speed, accuracy, or force. A fourth is in terms of risk or actual detriment to his health or future capabilities. The common technique of defining workload as the negative of "spare capacity" (objective performance on a "side task") would fit the third category. The greatest problems in workload analysis seemed to be in finding operational definitions which are common across a variety of tasks.

Internal Models of External Systems

Keen interest in "internal modeling" was evident. The internal model is as old as the ancients: the notion is that somehow the brain has an internal representation of structures in the external world - of which it can ask "what would happen if ... " and get reasonable answers. In more recent years, "plant-model" controllers and Kalman filters have become useful tools in control, and human operator modelers have hypothesized that people make use of such internal models too - at least that is a convenient normative base from which to model. For example, the well-known "Kleinman-Baron-Levison optimal control model" of the human operator includes a Kalman predictor as an internal model, and so do recent failure-detection models of Curry and Gai.

When considering man-computer cooperation in supervisory control systems, the problem arises whether the human operator needs an internal model not only of the process or vehicle to be controlled but also of the decision-making processes which involve the computer serving as a lower-level controller. Such a consideration elicited further concern by some Symposium participants that the computer needs not only an internal model of the process to do its control job, but must also have some model of what the human is likely or able to do in order that it can decide when to relinquish control to the human or take back control. More is said of this man-computer responsibility problem later.

Representation of Goals and Performance Criteria

Closely associated with the "internal models" discussions were those of performance criteria, objective functions, cost functions, payoff functions, tradeoff functions (and other terms, all of which mean the specification of what is good and what is bad for a particular system).

It is not easy for the experimenter to communicate to an experimental subject a precise performance criterion or objective function, nor can a human operator in an actual system explain precisely what function of relevant variables he is seeking to maximize. Attempts to communicate such functional relationships in mathematical form almost always fail because most people simply don't understand their own behavior in such terms. Thus the experimenter or the analyst of actual system behavior must infer the performance criterion actually being used - either from observation of actual system behavior or from a battery of subjective judgments which the subject says he would make if confronted with given particular situations.

If inference of performance criterion is made from subjective judgments,

there are several techniques which are under active development. One of these is "multidimensional scaling" - a technique developed at Bell Telephone Laboratories and based upon subjective judgment of dissimilarity between multi-dimensional stimuli. By least-squares regression it yields "principal axes" of difference between stimuli. A second technique is "multi-attribute utility", developed by Raiffa, Keeney and others, and based on indifference judgments of utility (or worth) with respect to lotteries. It yields multi-dimensional cardinal scales of utility. A third technique called "interpretive structural modeling" was developed by Warfield and his colleagues at Battelle Institute, and is appropriate where transitive orderings can be made among stimuli. A fourth technique called "policy capturing" makes use of direct worth assessment. And so on. All of these techniques have interactive computer procedures to accompany them.

Tradeoffs in System Strategy

The Symposium also dealt with what might be called "system strategy tradeoffs" - questions of management philosophy regarding overall strategy for making the system work best.

Perhaps two such tradeoffs are most important. The first is the question of relative power of the human operator vs. the computer over one another. Who monitors whom? Should the human operator always have the upper hand, the final word? Or should system designers endow the computer with the capability to overrule erratic and detrimental decisions of the operator? To what extent should man and computer be working in parallel, independently and in relative ignorance of what the other is intending? Or is it preferable that man and computer cooperate in close harmony, each being continually updated on what the other is doing and intending? These questions were in no way resolved. But the need for further research was evident.

A second "strategy tradeoff" is between acquiring enough data to be confident of action vs. acting with anticipation. This is a somewhat older problem of cybernetics, and does have a quantitative solution provided the event states are well defined, and all the appropriate prior and contingent probabilities of these events are known, as well as the costs/benefits of possible outcomes.

Tradeoffs in Modeling

Questions of modeling philosophy were of particular interest because, as systems become more complex and the role of the human operator becomes more "supervisory" and thus less amenable to definition, modeling becomes more difficult.

These questions might, again, be posed in terms of trading relations:

a) Mathematical models of man-machine systems are complex and incomprehensible to many people to begin with, and lately are becoming more so. Yet the more traditional verbal model of classical behavioral and social science is often ambiguous and not seen as particularly useful in doing an engineering job. Is there a practical compromise between mathematical complexity and verbal ambiguity?

b) Controlled experiments require a constraint and distortion of reality,

and the experimental subject knows this. Thus validity of his behavior is questioned. Reality, on the other hand, is chaotic. How to trade between the unreality of experimental control and the chaos of reality?

c) A common criticism of man-machine systems and technological "progress" generally is that the people involved are more and more forced to behave like (and be modeled like) machines. Should that trend continue? Or should there be a counter-trend to enable human operators to behave and be modeled as fully free, creative, social, emotional people?

d) As engineering areas have matured there have been efforts to standardize certain models, definitions, measures, etc. across the professional community. Should no standardization be introduced in this field, is it impossible or premature, or should some standardization efforts be pursued?

e) Man-machine modelers have been criticized as being "mechanistic", "logical-positivist", "behaviorist", and therefore presumably out of date, especially as we move toward supervisory control. The opposite might be a "gestalt", "holistic" approach. E.R.F.W. Crossman has used the term "polyvalent craftsman" to characterize the human supervisor. Can modeling in such terms be useful, and what compromises are appropriate?

Ethical Implications

There is much talk these days of making individuals, companies, and nations more "productive". Mostly, it seems, "productivity" is considered in functional terms - speed, reliability, etc. At the Symposium there was great concern, especially among the Scandinavian delegates, about productivity in human terms - job satisfaction- and whether work is meaningful.

What changes is the computer making with respect to job satisfaction, not only as regards employment statistics and training requirements? We are warned that the computer and associated automation may be making the human worker spatially more remote from actually handling the product, and temporally less synchronized with the rhythms of traditional work. The worker may perceive that he has less and less direct effect on the product, is more and more specialized, and has less and less understanding of the production process generally.

The rumblings in this problem area seemed a bit more significant than I remember them to have been in previous meetings.

HUMAN CONTROL AND MONITORING-MODELS AND EXPERIMENTS

By P.H. Wewerinke

National Aerospace Laboratory (NLR)

SUMMARY

This paper deals with the results of a theoretical and experimental program concerning human monitoring behavior. Apart from monitoring an automatic approach, combined monitoring and manual flight director control was studied to determine the interference between subtasks. Also simultaneously monitoring and auditory tracking was included.

The results demonstrate that the multivariable monitor model adequately describes human behavior in the aforementioned tasks. Furthermore, a multivariable workload model is developed. Computed workload is shown to agree excellently with subjective ratings.

INTRODUCTION

With increasing complexity and automation of aerospace vehicles the human operator's role shifts from controller to supervisor. This necessitates the proper tools for describing these manned vehicle systems. Apart from controlling the system (to which situation considerable modeling effort has been devoted), the pilot often has to fulfil other crucial functions, such as monitoring the automatic system, making decisions, detecting system failures, etc. The insight in this higher mental functioning is still rather incomplete although several attempts have been made to investigate and model signal- and failure detection behavior.

The object of this study was to describe human monitoring behavior and to determine how it is affected by performing other tasks (interference). For this, a model for monitoring multivariable systems has been developed which can be considered as an extension of the research of Levison et al (Ref. 4). Essentially, it is a cascade combination of the subjective expected utility model (Refs. 2 and 3) and the state estimation submodel of the optimal control model.

The model has been tested against experimental results of a fixed base simulator program, dealing with monitoring the automatic approach of a DC-8. Also combined monitoring and manual flight director control (using the optimal

control model) was investigated to determine the interference between several subtasks using the task interference model given in reference 5. For the same reason, the monitor tasks were combined with an auditory tracking task.

In section 2 the human operator models involved will be briefly reviewed. Section 3 contains a discussion of the experimental program and a comparison of the predicted and measured results. In section 4 pilot workload is analyzed and a multivariable workload model is developed in accordance with the task interference model and fractional workload model suggested in reference 5.

HUMAN OPERATOR MODELS

This section summarizes the mathematical models which are used to predict and analyze the results of the experimental program discussed in the next section. The models deal with human control-, monitor- and decision making behavior as well as their mutual interference.

Human control behavior is described by the optimal control model (OCM) developed by Kleinman and Baron (Ref. 1). This model consists of a perceptual model indicating how the "displayed" variables are related to the "perceived" variables, and a response model. The latter model describes how the internal representation of the task environment results in the actual control input(s).

Human decision making behavior is described by the perceptual model, however now in a cascade combination with the subjective expected utility model (Refs. 2-4). This model, reflecting how "perceived" information results in an optimal (binary) decision strategy, will be briefly discussed in the next chapter.

Interference between several tasks is accounted for by the task interference model developed and partly validated by Levison et al (Refs. 4 and 5). This model plays a major role in the present study not only to predict the interference between the pertinent control- and decision making tasks but also to formulate the fractional and total workload involved in performing the tasks. For this reason the task interference model is briefly discussed in chapter 2.2.

Human operator workload is expressed in terms of the fractional attention corresponding with each subtask (Ref. 5). One objective of the present study was to extend and validate this model in order to predict human operator workload corresponding with multivariable control- and decision making tasks.

Decision making model

Perceptual model

Decision making is assumed to be based on the internal representation of the state of the world, \hat{x} . This internal model of the system state, x , is based on the "perceived" variables, y_p according to

$$y_p(t) = y(t-\tau) + v_y(t-\tau) \quad (1)$$

where y represents the "displayed" variables, τ is a lumped "equivalent" perceptual time delay, and v_y is a vector of independent, white, Gaussian observation noises. The autocovariance of each noise component appears to vary proportionally with mean-squared signal level and may be represented as

$$v_{y_i}(t) = \tau P_{y_i} E \left\{ y_i^2(t) \right\} \quad (2)$$

where P_{y_i} is the "noise/signal" ratio corresponding with the fractional attention paid to variable y_i (see chapter 2.2).

Denoting the covariance of the difference between the instantaneous value of the state vector and the estimate of it ($e(t) = x(t) - \hat{x}(t)$) by Σ , the pair (\hat{x}, Σ) constitutes a sufficient statistic to test hypotheses about x , based on the data y_p .

Subjective expected utility model

It is assumed that the human's decision strategy is reflected by the following stages (Refs. 2-4).

- . formulate (N) possible hypotheses, H_j
- . assess probabilities of all hypotheses based on the available information y_p , $P(H_j/y_p)$
- . determine (M) possible decisions, D_i
- . assign the utilities to each hypothesis/decision combination U_{ij}
- . determine the maximum utility-decision D^* according to $D_i = D^*$ for $E = E_{\max}$, where

$$E \{ U/D_i \} = \sum_{j=1}^N U_{ij} P(H_j/y_p) \quad (3)$$

Many decision making situations involve binary decisions (the choice between two hypotheses, e.g., the probability of a successful landing is large enough or not). Both the following analysis and experimental program deal with this class of decision tasks.

Now, let the mutually exclusive hypotheses H_0 and H_1 correspond to the events that $x \in R$ and $x \notin R$, respectively, where R is defined as some region (or window) in the state space. Using eq. (3) the human's decision is therefore given by

$$D = D_1 \text{ if } \frac{P(H_1/y_p)}{P(H_0/y_p)} > \frac{U_{00}-U_{10}}{U_{11}-U_{01}} \equiv \frac{U_0}{U_1} \quad (4)$$

$$D = D_0 \text{ otherwise}$$

In the following the target region R is given by a multidimensional "window"; then for H_0 can be written

$$H_0 : x_{TL_i} < x_i < x_{TU_i} \text{ for all } i, i=1, \dots, n \quad (5)$$

where x_{TL_i} and x_{TU_i} represent the lower and upper target boundary, respectively. Computation of eq. (4) requires the posterior probability of hypothesis H_0

$$P(H_0/y_p) = \int_{x_{TL_1}}^{x_{TU_1}} \dots \int_{x_{TL_n}}^{x_{TU_n}} p(x/y_p) dx_1 \dots dx_n \quad (6)$$

where $p(x/y_p)$ is (assumed to be) Gaussian with mean \hat{x} (maximum likelihood estimate) and covariance Σ . For the computation of various measures of decision performance it is convenient to define the human's decision space DS using eq. (4)

$$\frac{P(H_1/y_p)}{P(H_0/y_p)} = \frac{U_0}{U_1} \quad (7)$$

or

$$P(H_0/y_p) = \frac{U_1}{U_0 + U_1} \quad (8)$$

Combining eqs. (6) and (8) yields

$$\int_{x_{TL_1}}^{x_{TU_1}} \dots \int_{x_{TL_n}}^{x_{TU_n}} p(\delta, \Sigma) dx_1 \dots dx_n = \frac{U_1}{U_0 + U_1} \quad (9)$$

where $p(\delta, \Sigma)$ represents the normal probability density function with mean δ (representing the elements of DS) and covariance Σ . Now the probability that decision D_0 is made can be computed according to

$$P(D_0) = P(\hat{x} \in DS) \quad (10)$$

The probability of wrongly deciding D_0 is given by

$$P(H_1, D_0) = P(x \notin R, \hat{x} \in DS) \quad (11)$$

Alternatively, the probability of wrongly deciding D_1 can be computed according to

$$P(H_0, D_1) = P(x \in R, \hat{x} \notin DS) \quad (12)$$

Foregoing theoretical decision making measures will be compared with the corresponding measures obtained in the experimental program discussed in section 3.

Task interference model

Task interference is modeled in terms of the following relationship between fraction of attention, f_i , paid to subtask (indicator) i , and the corresponding human's internal noise/signal ratio, P_i

$$P_i = P_{0_i} / f_i \quad (13)$$

where P_{0_i} is the ratio corresponding to single-task performance ("full attention"). Furthermore, it is assumed that the amount of information-processing capacity is determined by the demand of the subtasks and not by the amount of subtasks to perform. In formula

$$\sum_{i=1}^M f_i = 1 \quad (14)$$

This model developed by Levison et al, has been partly validated for some multivariable control situations (Ref. 5) and dual decision making tasks (Ref. 4). In the present study the model will be tested in multivariable

hypotheses-situations, and in combined control and decision making tasks (both interacting and non-interacting).

In reference 5 the concept of fractional attention, is related to fractional workload. Based on this suggestion a multivariable workload model (both for control and decision making situations) will be presented and compared with experimental data (subjective ratings). The model will be discussed in section 4, because it has not been used to predict the experimental results.

EXPERIMENTS

Experimental set-up

In order to investigate monitoring behavior for various task situations, the following single-task configurations were examined in the experimental program (figure 1)

- . monitoring the fast/slow- and the glideslope indicator during an automatic approach of a DC-8 (indicated by M2).
- . monitoring the fast/slow-, the glideslope- and the localizer indicator (M3).
- . manual (flight director) approach (only longitudinally) of a DC-8 (C).
- . auditory presented unstable first order tracking task with a time constant of 1 sec (A).

Also combinations of these tasks were included to validate the aforementioned task interference model for both control and monitoring tasks. Combining the longitudinal control task and the two monitoring tasks results in "two" interacting tasks, or put another way, the prior probabilities of the monitor tasks depend on the human control behavior. The combinations of the monitor tasks and the auditory tracking task were included to study the interference between the monitor tasks and non-interacting (side-) tasks (e.g., radio-communication, procedural tasks, etc). So, the resulting combined task configurations are M2C, M3C, M2A and M3A.

Single-task configurations

Referring to figure 1, the decision making tasks were intended to represent the pilot's task of deciding whether or not he is within the "landing

window". Each indicator was displayed along with two reference indicators showing the "target" or region of acceptable deviations. The subject depressed a response button whenever he decided hypothesis H_1 to be true (one or more indicators outside their region of acceptance).

The longitudinal control task was a manual approach tracking task (without the time varying aspect involved in an actual approach and landing task) using a flight director (FD in figure 1). The flight director design, autopilot characteristics and turbulence levels are derived from reference 6.

The first order instability task was presented auditory. The display characteristics were such that the perceived tone (pitch) was linearly related to the system output (Ref. 7). No external driving noise was included, so the subject's task was to minimize his remnant by manipulating an isometric side-arm controller.

Measures, subjects and procedures

For the decision making tasks the measures discussed in section 2 were taken. Tracking performance was measured in terms of the relevant variance-scores (ILS-deviations, flight director deviations, speed deviations, stick activity, and audio display-deviation). Furthermore, pilot workload was measured in terms of subjective ratings on the scales presented in table 1. In case of the dual task configurations, an overall impression of the total task difficulty was given.

Four general aviation pilots participated in the experiment. The subjects were provided with about fifty training trials totally (six on each of the eight configurations) corresponding with a relatively stable level of performance. In the formal experiment each configuration was presented four times per subject. The duration of the trials was five minutes and the order of presentation of the trials per subject was random. In the combined control- and decision making task the subjects were instructed to perform the control task as well as possible and to spend their "reserve capacity" on the decision making task.

Theoretical and experimental results

Since space does not permit an extensive presentation of the experimental results*, only the principal results will be discussed corresponding with the model predictions.

* These will be contained in a later report.

Single-indicator tasks

In order to obtain noise/signal ratios corresponding to monitor single-display indicators ("full attention") a base-line experiment was conducted to "calibrate" the displays. Three subjects performed each of the three single-indicator decision making tasks two times. The measured decision making scores are shown in table 2. Two model parameters were varied to match the corresponding scores: the observation noise level associated with the displayed variable and the utility ratio U_1/U_0 . The perceptual time delay, τ , was kept constant (0.20 sec). The resulting model scores are also given in table 2. Although the subjects were instructed to weigh "miss"-errors and "false alarm"-errors equally and to minimize the total decision error, P_e , they were apparently somewhat reluctantly in deciding "in" ($U_1/U_0=1.75$). Based on this result $U_1/U_0=1.5$ was used for the subsequent model predictions. The resulting observation noise levels* reflect display phenomena and/or "indifference" thresholds. These values were assumed to be the ratios corresponding to "full attention" (P_{oi}).

Using eq. (13) the actual noise/signal ratio could be determined given the fraction of attention, f_i , paid to indicator i for the multivariable decision making (and control) tasks.

Multivariable decision tasks

Decision scores were predicted for the M2- and M3-tasks assuming an equal division of attention among the display indicators. A comparison of measured and predicted scores (Table 3) reveals an excellent agreement between all corresponding scores. The effect of the ratio U_1/U_0 on the decision making scores is shown in figure 2. As expected the probability of deciding "in" as well as the joint probabilities are relatively sensitive to this ratio; however, the total decision error, P_e , increases only slightly when U_1/U_0 varies from 1 to 1.5**.

Decision making and non-interacting control

Next the combined auditory tracking and decision making tasks are considered. The following procedure was used.

* Observation noise levels corresponding with indicator position are modified because of the non-zero references (Ref. 4).

** The assumed value of 1.5 turned out to be rather close to the actually "measured" ratio. Pilot's comments clarified why this ratio was larger than one: after being out the "window" they wanted to be "sure" before deciding "in" again.

The noise levels of the single audio task were determined by matching the model scores to the measured ones using figure 3; these values correspond to "full attention". The same was done for the audio tasks when performed in combination with the decision tasks. The corresponding fractions of attention (f_A) were used to predict the attention paid to the decision tasks ($1-f_A$). The resulting decision scores are given in table 3. Comparing these scores with the measured errors shows that the model predicts a larger interference than actually measured (about ten percent too large for both M2(A) and M3(A)). Matching the measured decision scores using figures 4 and 5 containing the theoretical curves of decision error versus fraction of attention results in an actual fraction of attention paid to M2(A) of 0.7 and to M3(A) of 0.73.

The explanation for the smaller interference is, of course, the fact that the decision task involves visual information and that the control task is presented auditory. So, we see that the interference between visual and aural information is less than between two visual tasks. It is particularly interesting to note that the amount of interference for the combined M2(A) task is just the same as for the M3(A) task: the total amount of information processing capacity is for both configurations 1.2. Based on these results it is concluded that the task interference model has to be adjusted for multi-modality tasks. For combined visual and auditory tasks the total amount of information processing capacity is tentatively hypothesized to be 1.2 (instead of 1 (2) for full (no) interference).

Decision making and interacting control

The third set of configurations was included to determine the extent to which the task interference model would hold for simultaneous control and decision making when control behavior affects the (a priori) decision making statistics. For this reason the aforementioned split-axis manual approach and monitor task was investigated. The basic assumption in modeling this situation was that control behavior was based on the perception of the flight director and decision making on the basis of the perceived monitor indicators (u, d and y). In other words, it was assumed that two separate internal models were used.

The same procedure was followed as before. The single control task (C) was modeled by matching the model scores to the measured ones. The resulting correspondance (shown in table 4) is obtained for a noise ratio of -18.7 dB. By varying only the observation noise ratio the control tasks when performed in combination with the decision tasks (C(M2) and C(M3)) were matched. The result is also contained in table 4. The corresponding fractions of attention (f_C), shown in figure 6, were used to predict the attention paid to the decision tasks ($1-f_C$). The resulting decision scores are given in table 3. Comparing these with the measured errors shows an excellent agreement for the M3(C) task, however, the predicted errors for the M2(C) task are considerably larger (20 %) than the measured errors. For this reason the measured decision scores were matched using figure 7. The resulting scores (given in

table 3) correspond with a fraction of attention which is about 0.1 smaller than the predicted fraction (based on the interference model).

It is concluded from the foregoing results that the total information processing capacity is relatively constant (1.0 for the M3(C) task and 0.9 for the M2(C) task). However, an interesting refinement to the model will be discussed in the next section incorporating the aspect of pilot workload.

PILOT WORKLOAD

In accordance with the task interference model and the fractional workload model suggested in reference 5, a multi-task workload model is proposed in this section and compared with measured workload (subjective ratings).

Assuming the workload corresponding with performing a single task (indicator), W_o , the fractional workload when performing this task in combination with other tasks is

$$W_i = W_o F(f_i) \quad (15)$$

where $F(f_i)$ represents the functional relationship between the fraction of attention¹ dedicated to the subtask i and the corresponding fractional workload. The total workload involved in performing M tasks is given by

$$W_t = \sum_{i=1}^M W_i \quad (16)$$

implicitly assuming that the function F is such that W_i is expressed in units along an interval scale.

Subjective ratings were used which were obtained on the adjectival "demand-scale" shown in table 1 (Ref. 8). This scale is assumed to be an interval scale. Assuming a logarithmic relationship between f_i and W_i results in $F(f_i) = \text{Log } cf_i$ where c is an empirical constant corresponding to the zero of the interval scale. A value of $c = 10.5$ results in computed total workload values surprisingly close to the measured workload ratings. This is shown in table 5 and plotted in figure 8*.

Returning to configuration M2(C) one can see that the computed workload based on the predicted fraction of attention on the monitor task is larger than the measured workload. This is in accordance with the predicted performance which was superior to the measured decision error. Using the "matched"

* During the single-indicator experiment no workload data were obtained. Therefore, W_o , etc. were assumed to be equal and estimated by matching the M2 configuration. The result was a rating of 3.1.

fraction of attention results in a predicted workload very close to the value actually measured (see table 5 and figure 8). So, the only question is why the predicted performance-workload trade-off differs from the measured one.

In figure 9 the theoretical curves of decision error versus fraction of attention are shown for both M2(C) and M3(C). For the matched M2(C) task the performance versus workload trade-off, $\Delta P / \Delta W = .12$. This is close to the value for the M3(C) configurations (.11).^e So, the suggestion is that the subjects were not motivated enough to spend their "full" capacity because the pay-off was insufficient. More experimentation will be needed to establish this capacity/pay-off relationship. However, this can be considered as a (useful) refinement to the models which have been shown to describe encouragingly the control and decision making tasks studied in this program, both in terms of performance and workload.

REFERENCES

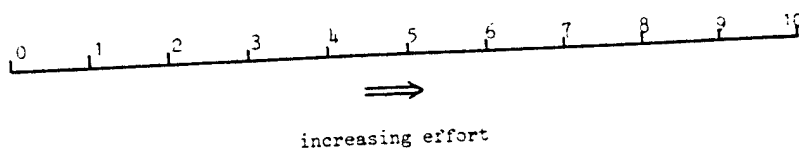
1. Kleinman, D.L. and Baron, S.: Manned Vehicle System Analysis by Means of Modern Control Theory. NASA CR-1753, June 1971.
2. Edwards, W. et al: "Emerging Technologies for Making Decisions", in New Directions in Psychology II. Holt, Rinehart and Winston, New York, 1965.
3. Edwards, W. and Tversky, A. (Ed.): Decision Making. Penguin Modern Psychology UPS 8, 1967.
4. Levison, W.H. and Tanner, R.B.: A Control-Theory Model for Human Decision Making. NASA CR-1953, December 1971.
5. Levison, W.H. et al: Studies of Multivariable Manual Control Systems: A Model for Task Interference. NASA CR-1746, May 1971.
6. Graham, D. et al: Investigation of Measuring System Requirements for Instrument Low Visibility Approach. AFFDL-TR-70-102, February 1971.
7. Wewerinke, P.H.: A Simulator Experiment to Investigate a Lateral Rate Field Display. NLR TR 74093, June 1974.
8. McDonnell, J.D.: Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities. AFFDL-TR-68-76, December 1968.

TABLE 1: RATING SCALES

Name:

Task:

Using the scale below, indicate the degree of effort you spend on performing the task



Rating scale for
Controllability/observability and Precision

- 0 - Extremely easy to control/monitor with excellent precision
- 1 -
- 2 - Very easy to control/monitor with good precision
- 3 -
- 4 - Easy to control/monitor with fair precision
- 5 -
- 6 - { Controllable/observable with somewhat inadequate precision
- 7 - { Controllable/observable, but only very imprecisely
- 8 - Difficult to control/monitor
- 9 - Very difficult to control/monitor
- 10 - Nearly uncontrollable/unobservable
- 10 ☐ Uncontrollable/unobservable
- ☐ Not applicable

Rating Scale for
Demands on Pilot

- 0 -
- 1 -
- 2 - { Completely undemanding, very relaxed and comfortable
- 3 -
- 4 - Largely undemanding relaxed
- 5 -
- 6 - { Mildly demanding of pilot attention, skill, or effort
- 7 - { Demanding of pilot attention, skill, or effort
- 8 - { Very demanding of pilot attention, skill, or effort
- 9 - { Completely demanding of pilot attention, skill, or effort
- 10 - Nearly uncontrollable/unobservable
- 10 ☐ Uncontrollable/unobservable
- ☐ Not applicable

DISPLAY INDICATOR	PROBABILITY	$P(H_0)$	$P(D_0)$	$P(H_1, D_0)$	$P(H_0, D_1)$	P_e	OBSERV. NOISE (dB)	
							$P_{\dot{y}}$	P_y
u	MEASURED	.70	.68	.02	.04	.06		
	MODEL	.70	.68	.02	.04	.06	-15.	-18.8
d	MEASURED	.82	.79	.02	.04	.06		
	MODEL	.82	.80	.02	.04	.06	-13.	-15.5
y	MEASURED	.75	.73	.03	.04	.07		
	MODEL	.75	.73	.03	.04	.07	-13.	-15.5

AVERAGES OF THREE SUBJECTS, 2 RUNS PER SUBJECT, $U_1^U = 1.75$

TABLE 2: COMPARISON OF MEASURED AND MODEL PROBABILITIES FOR THE SINGLE-INDICATOR TASKS

CONF.	PROBABILITY	$P(H_0)$	$P(D_0)$	$P(H_1, D_0)$	$P(H_0, D_1)$	P_e
M2	MEASURED	.56	.53	.050	.079	.129
	MODEL	.56	.52	.048	.082	.130
M2(A)	MEASURED	.56	.53	.057	.086	.144
	MODEL PREDICTED	.56	.52	.061	.098	.158
	MODEL MATCHED	.56	.52	.053	.089	.142
	MEASURED	.47	.43	.088	.116	.204
M2(C)	MODEL PREDICTED	.47	.41	.057	.110	.167
	MODEL MATCHED	.47	.41	.069	.132	.201
	MEASURED	.47	.44	.060	.088	.148
M3	MODEL	.47	.42	.053	.097	.150
	MEASURED	.47	.43	.065	.096	.161
M3(A)	MODEL PREDICTED	.47	.42	.067	.117	.183
	MODEL MATCHED	.46	.42	.060	.101	.161
	MEASURED	.36	.32	.086	.115	.201
M3(C)	MODEL	.36	.31	.072	.128	.201

AVERAGES OF 4 SUBJECTS, 4 RUNS PER SUBJECT

TABLE 3: COMPARISON OF PREDICTED AND MEASURED MONITOR PERFORMANCE

CONF.	SCORE	σ_u FT/SEC	σ_d FT	σ_{FD} DEG.DISPLAY	σ_{δ_e} DEG.
C	MEASURED	3.5	15.7	3.5	2.0
	MODEL	3.4	15.7	3.7	1.9
C(M2)	MEASURED	3.7	17.0	4.0	2.3
	MODEL	3.6	17.0	4.2	2.0
C(M3)	MEASURED	3.7	18.4	4.3	2.2
	MODEL	3.7	18.3	4.5	2.1

AVERAGES OF 4 SUBJECTS, 4 RUNS PER SUBJECT

TABLE 4: LONGITUDINAL CONTROL TASK - COMPARISON OF MEASURED AND MODEL SCORES

CONF.	SUBTASK	W_o	f_i	$F(f_i)$	W_i	W_t	MEASURED WORKLOAD
M2	M_u	3.1	.50	.72	2.2	4.4	4.4
	M_d	3.1	.50	.72	2.2		
M2A	A	3.4	.50	.72	2.4	5.6	5.7
	M2	4.4	.50	.72	3.2		
M2C	C	5.8	.43	.66	3.8	7.2	6.7
	M2	4.4	.57	.78	3.4		
	C	5.8	.43	.66	3.8	6.8	
	M2	4.4	.46	.68	3.0		
	M2	4.4	.67	.85	3.7		
M3	M_y	3.1	.33	.54	1.7	5.4	5.4
	A	3.4	.46	.68	2.3		
M3A	M3	5.4	.54	.75	4.1	6.4	6.5
	C	5.8	.25	.42	2.4		
M3C	M3	5.4	.75	.90	4.9	7.3	7.5

TABLE 5: COMPARISON OF COMPUTED AND MEASURED WORKLOAD

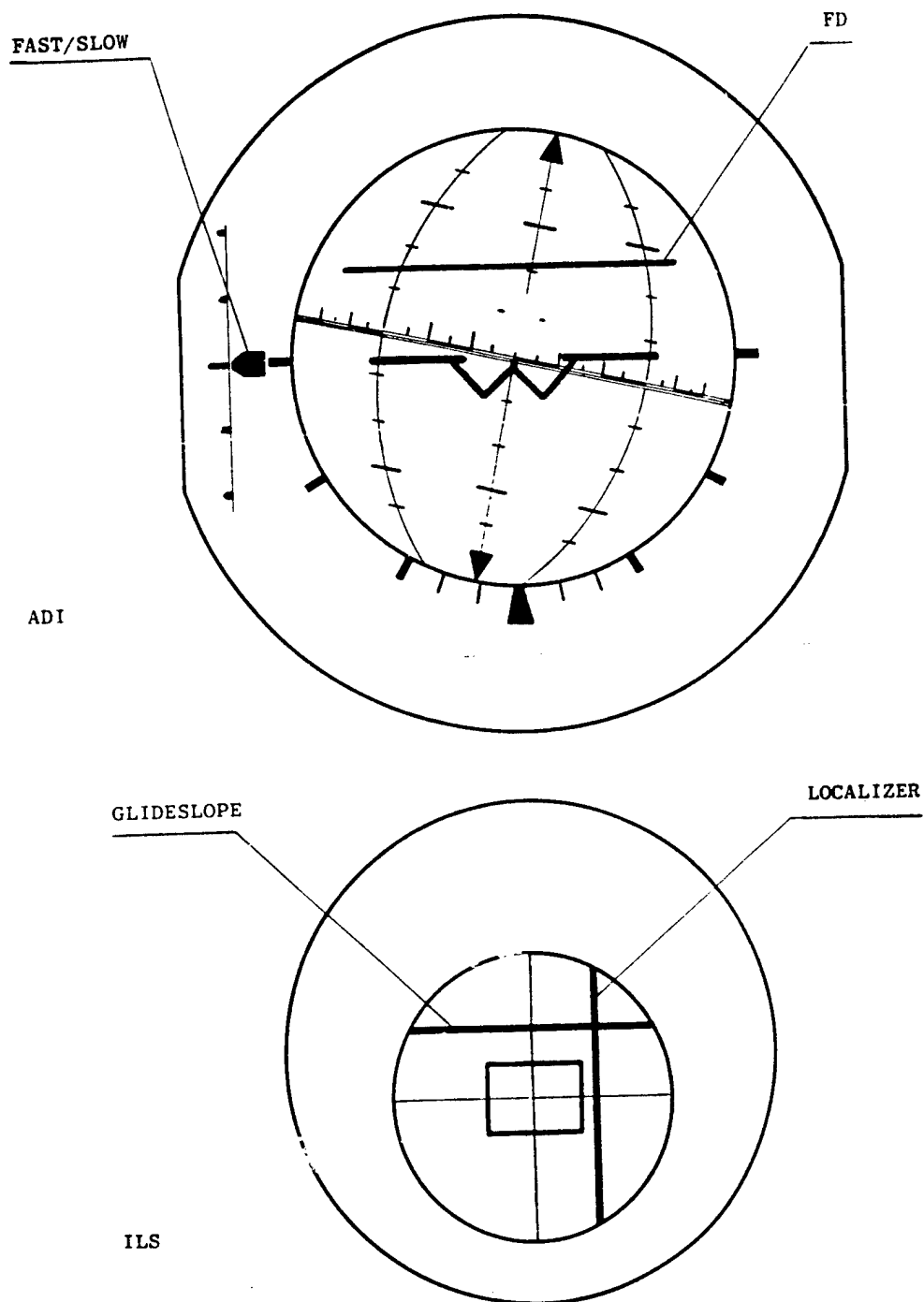


FIG. 1: DISPLAY FORMAT

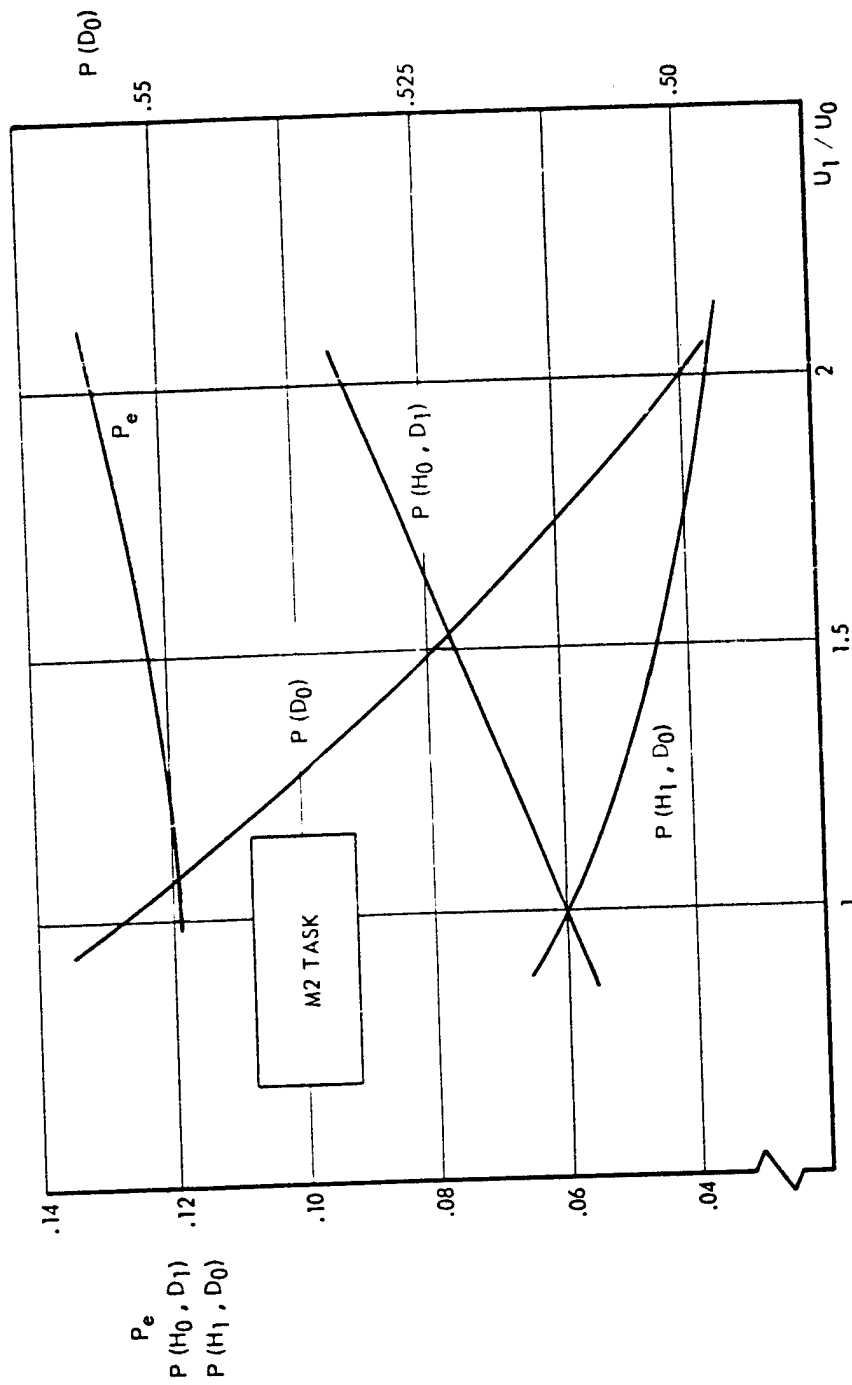


FIG. 2: THE EFFECT OF UTILITIES ON THE MONITOR PERFORMANCE MEASURES

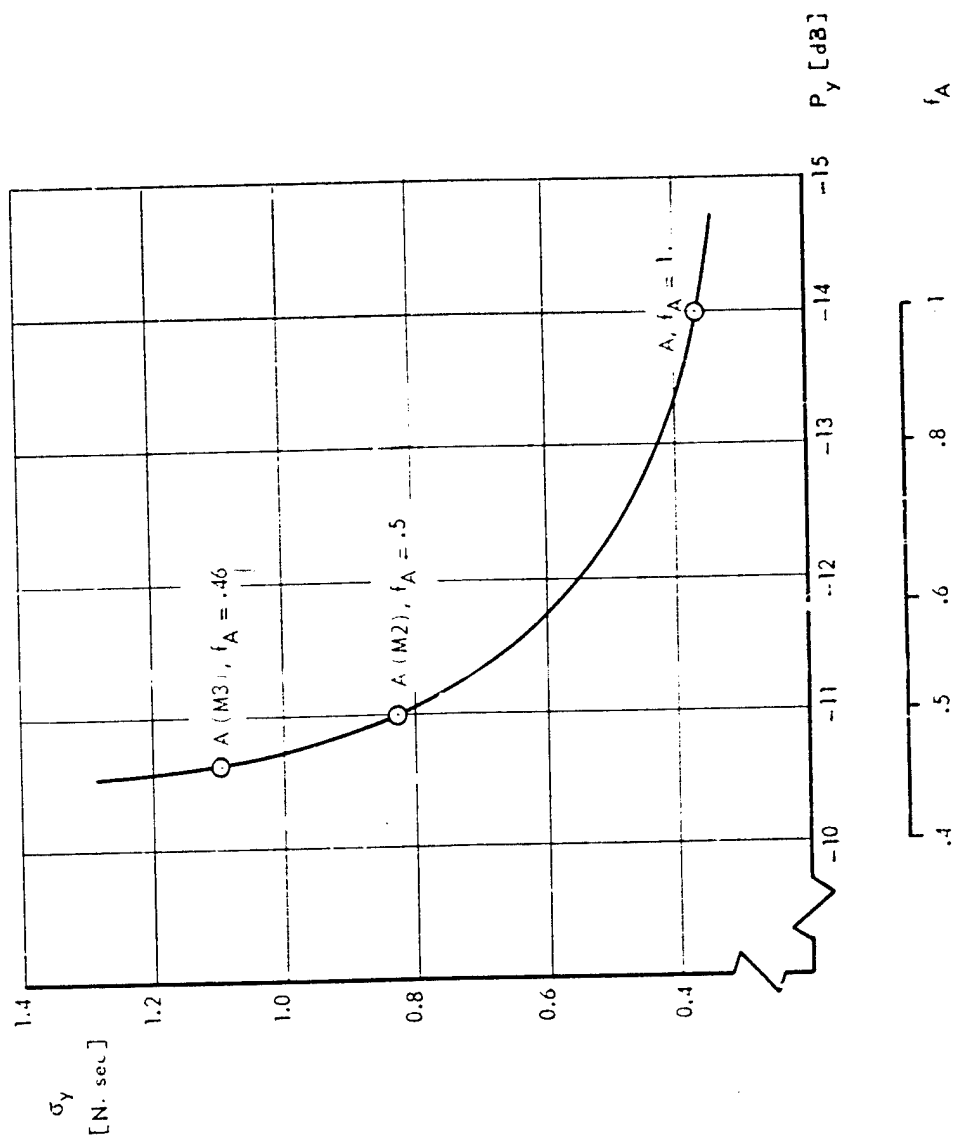


FIG. 3: AUDITORY TRACKING TASK - PERFORMANCE VERSUS FRACTION OF ATTENTION

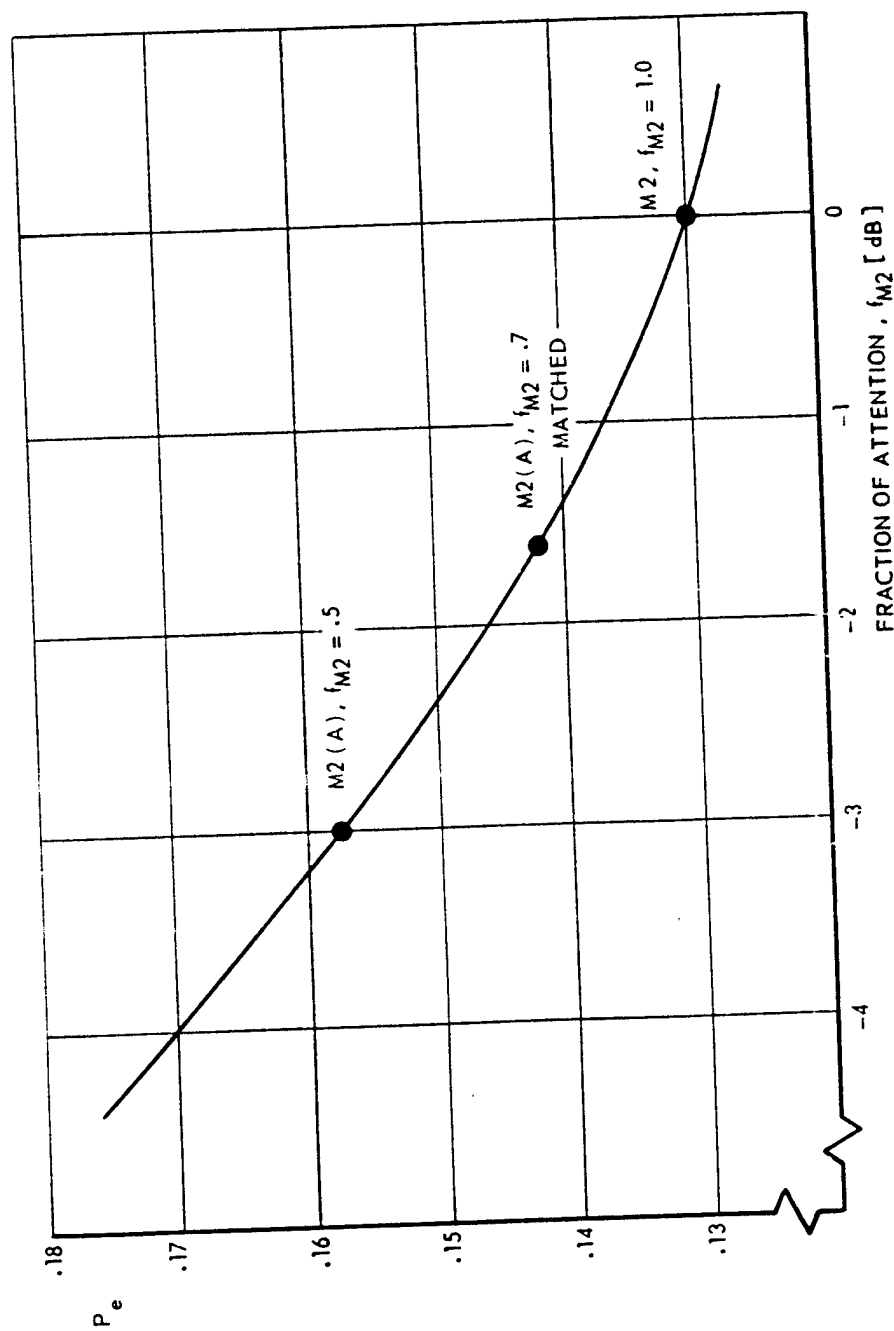


FIG. 4: MONITOR TASK, M2 - PERFORMANCE VERSUS FRACTION OF ATTENTION

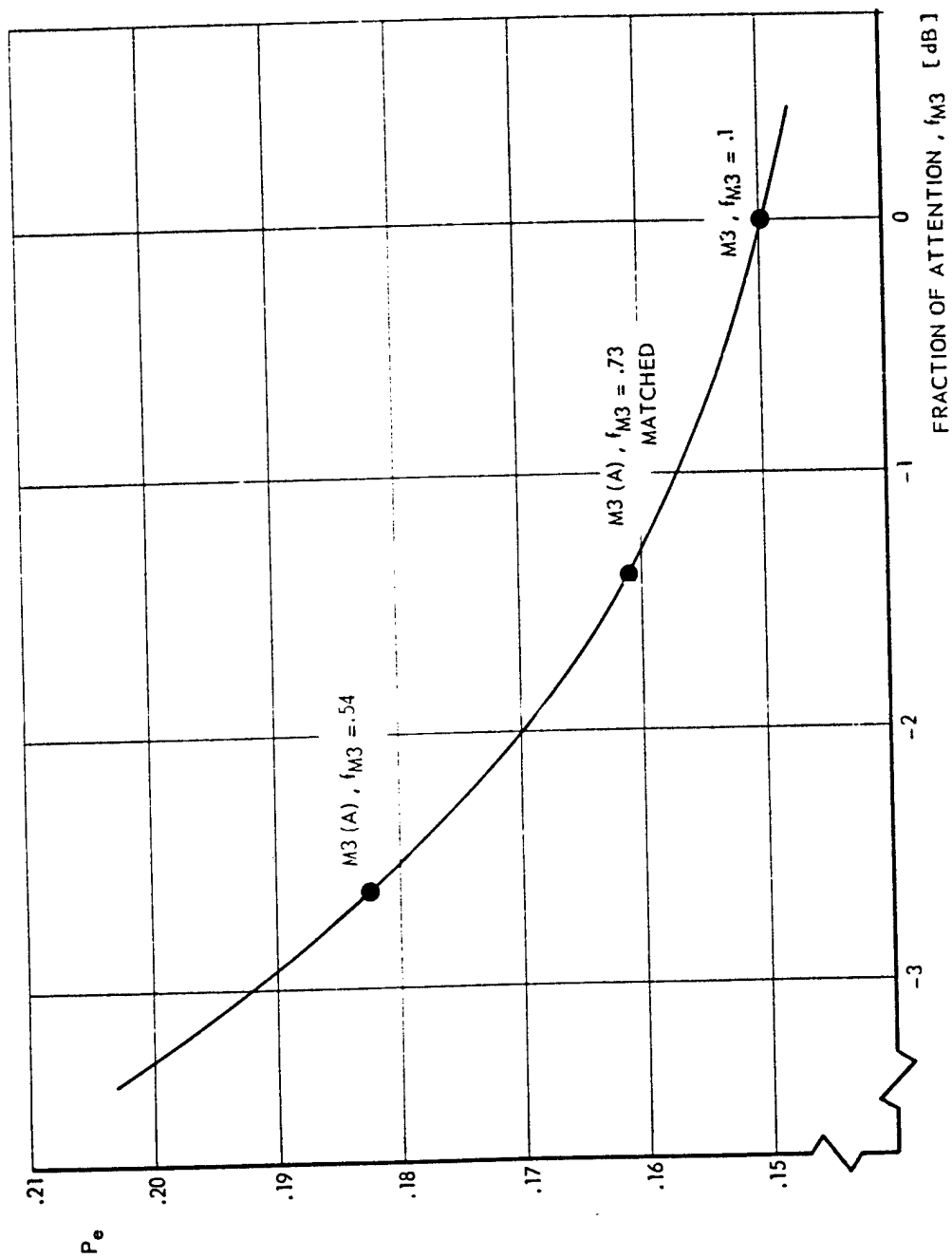


FIG. 5: MONITOR TASK, M3 - PERFORMANCE VERSUS FRACTION OF ATTENTION

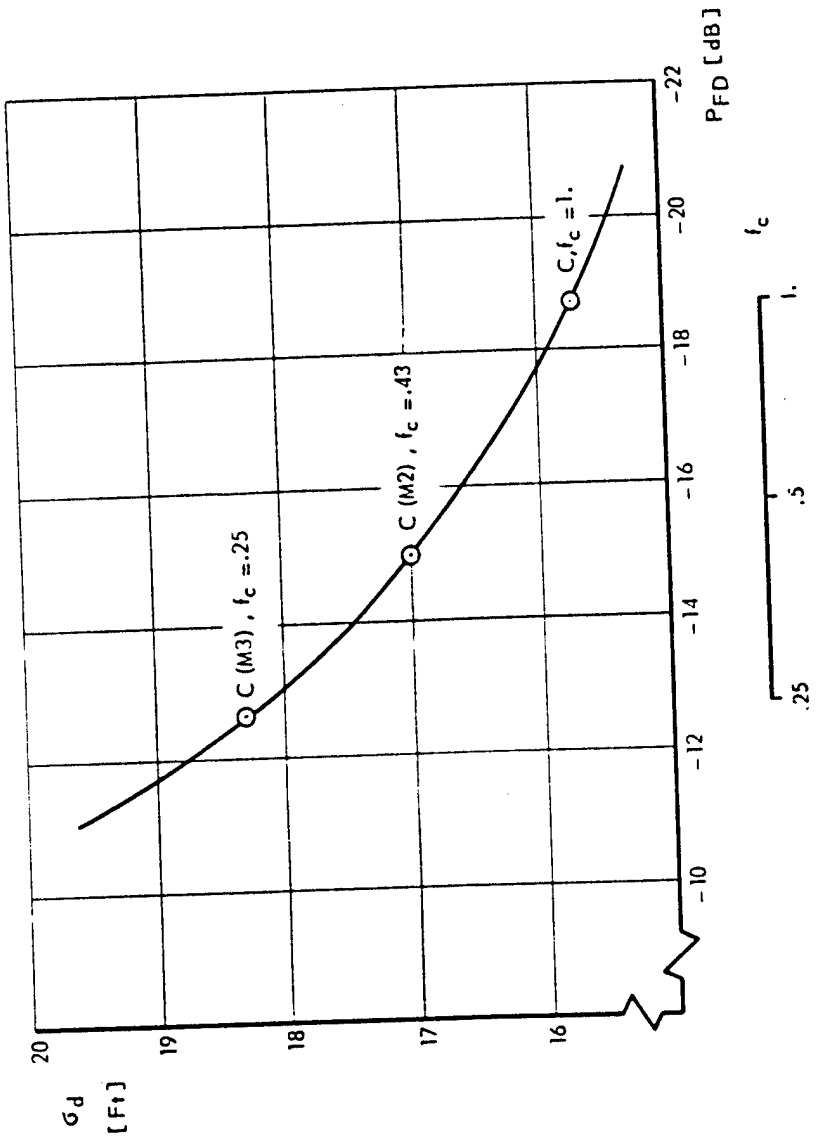


FIG. 6: LONGITUDINAL CONTROL TASK - PERFORMANCE VERSUS FRACTION OF ATTENTION

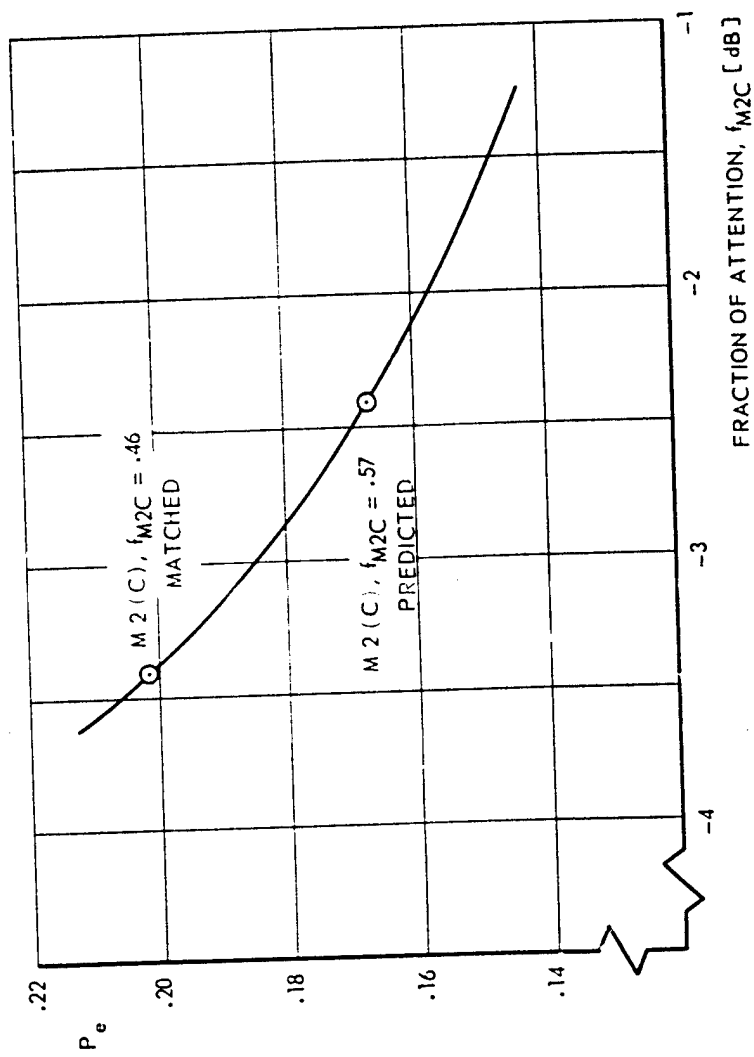


FIG. 7: MONITOR TASK, $M2(C)$ - PERFORMANCE VERSUS FRACTION OF ATTENTION

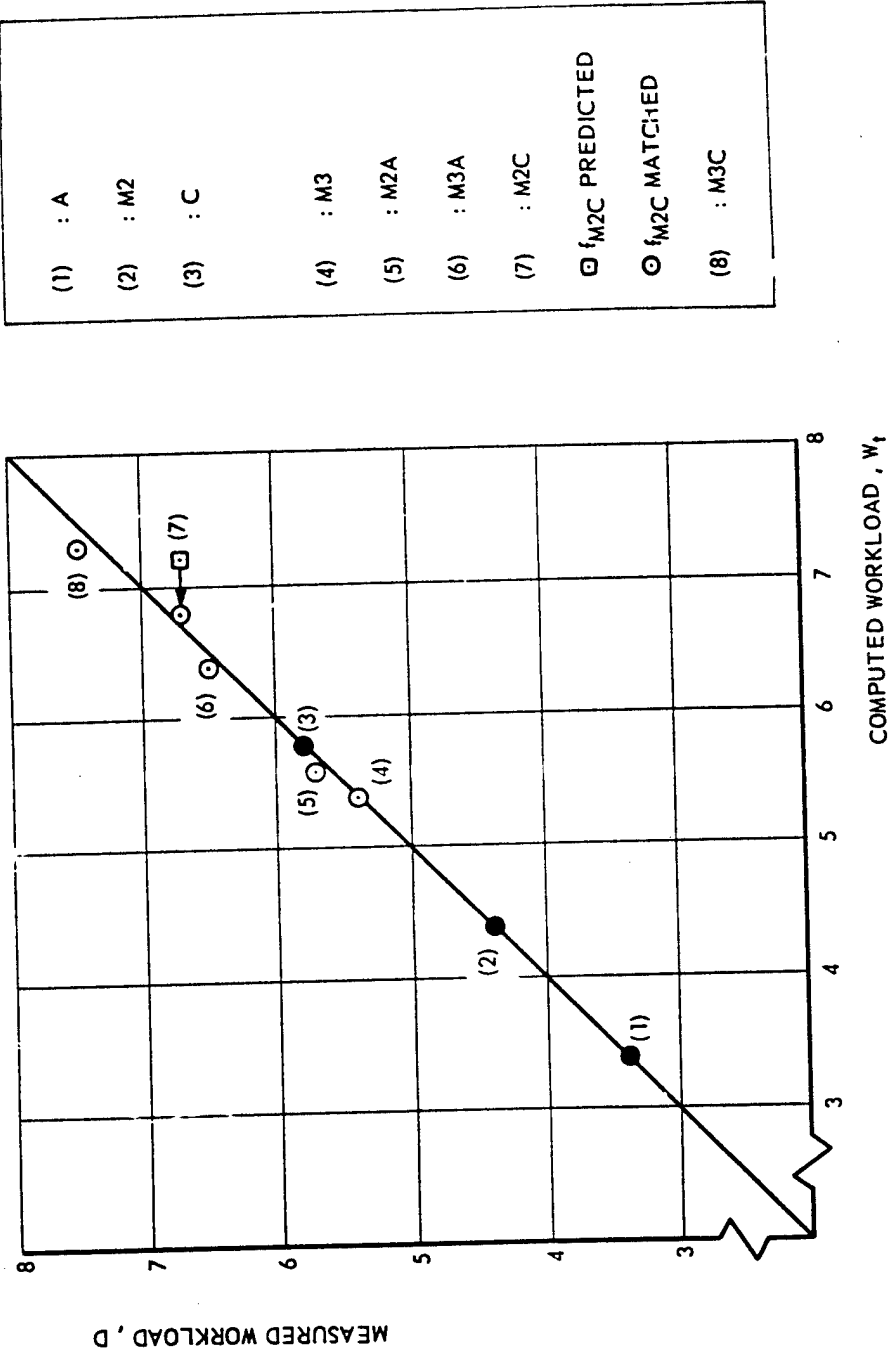


FIG. 8: COMPARISON OF MEASURED AND COMPUTED WORKLOAD

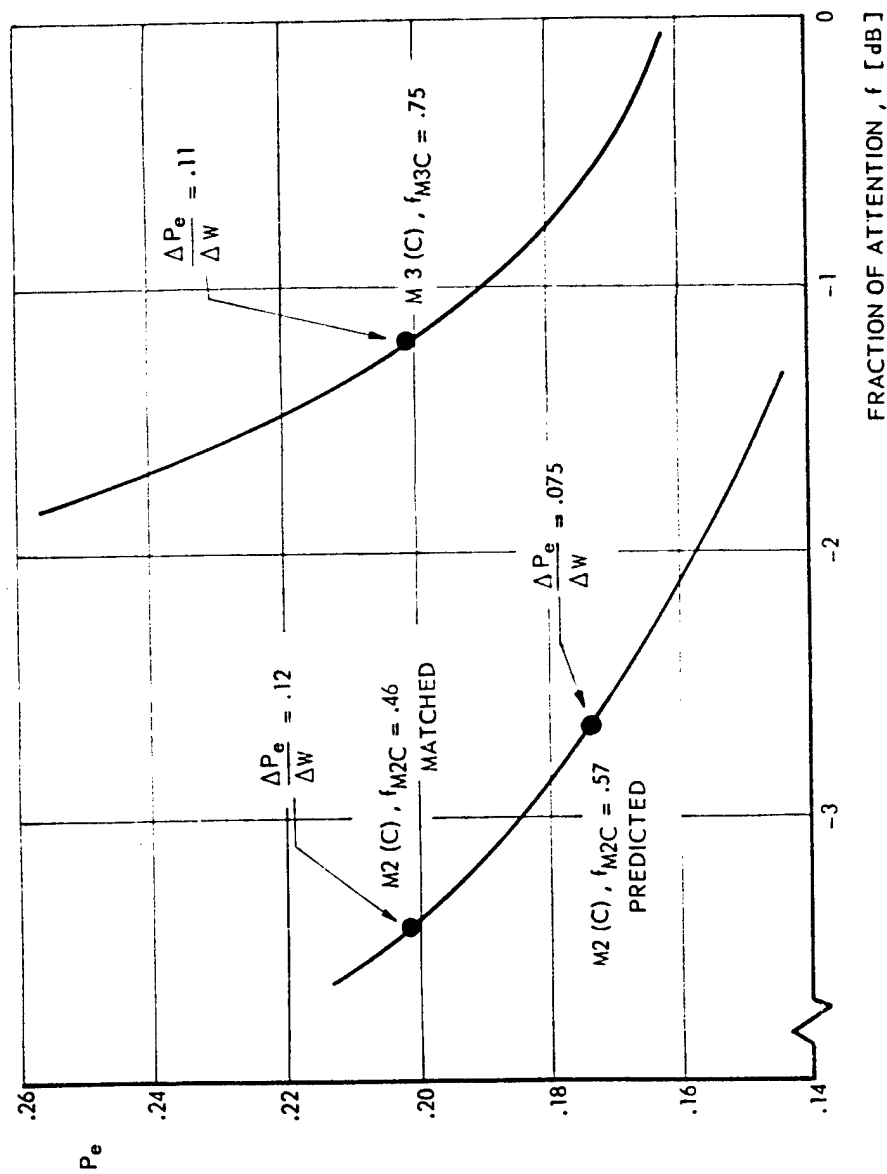


FIG. 9: MONITOR TASKS, MC - PERFORMANCE VERSUS ATTENTION AND WORKLOAD

AN EXPERIMENTAL SITUATION FOR STUDY OF PILOT INTERACTION
WITH AUTOMATED AIRBORNE DECISION MAKING SYSTEMS*

By William B. Rouse, Yee-Yeen Chu, and Rex S. Walden

University of Illinois
Urbana, Illinois 61801

SUMMARY

This paper describes an experimental situation which allows varying degrees of automation in a set of control and monitoring tasks. The control tasks involve flying a map display at specified altitude and speed. The monitoring tasks involve detection of events in N dynamic processes each of which has an associated display. Events evoke actions which involve a checklist-like tree search. Computer aid, with adjustable reliability, is available for both the control and monitoring tasks. A series of experiments which are to be performed in this experimental situation are described.

INTRODUCTION

Several factors are leading to the consideration of automated decision making systems for aircraft operations. Aircraft are becoming more sophisticated and complicated while greater precision and performance is being required of the pilot. The pilot simply does not have the time to do everything well. At the same time, computers are becoming smaller, faster, and cheaper. Thus, computer-aided decision making is (or will be) both desirable and feasible.

However, it is unlikely that the computer will completely replace the pilot. In failed or unusual situations, the pilot will be called upon to manually perform tasks normally allocated to the computer. Also, the pilot will serve as an executive or manager providing goals to the computer, monitoring overall performance, and occasionally preempting inappropriate decisions by the computer.

The problem area addressed by the project discussed in this paper is the interaction between a human pilot and a computer with decision making responsibility. The goal of the research is to enhance cooperation between the two decision makers by understanding and then avoiding possible modes of competition between them.

* Supported by the National Aeronautics and Space Administration under NASA - Ames Grant NSG-2119.

To be more specific, we are concerned with multi-task situations and the issues of primary interest include:

1. Allocation of responsibility between human and computer.
2. Resolution of conflicts between human and computer.
3. Human confidence in the computer system.

Considering allocation of decision making responsibility, we have proposed that responsibilities not be strictly assigned to each decision maker [1,2]. Instead, allocation should adapt to the state of the aircraft and the state of the pilot. Both pilot and computer would have responsibility for all or most decision making tasks with responsibility at any particular instant being assigned to the decision maker most able at that moment to perform the task. Simple concepts from queueing theory indicate that such a procedure for allocation of responsibility would improve the utilization of system (aircraft plus pilot) resources and thereby improve system performance. This approach would allow the pilot to retain a coherent role in the sense of having overall responsibility for the whole aircraft while the computer would enable the pilot to avoid having to continually exercise all of these responsibilities.

The most significant disadvantage of adaptive allocation is the possibility of conflict between the two decision makers. Without sufficient information about each other's actions, the pilot and computer might compete to perform tasks. This would degrade system performance (and perhaps be disastrous) and possibly cause adaptive allocation to be an unattractive approach. Theoretical and experimental approaches are being employed to assess the costs of conflict and devise methods of avoiding conflicts.

Of course, the success of any adaptive decision making system will depend on the pilot's confidence in the system. Within our experiments, we plan to study the pilot's decisions concerning mode of use of the computer system and his ability to detect when the decision making system has failed.

THEORETICAL FORMULATION

To investigate the feasibility of adaptively allocating decision making responsibility between human and computer, a mathematical model of multi-task decision making was developed and several simulation experiments were performed [3]. The goal of this effort was determination of the effects of several system variables including number of tasks, mean time between arrivals of action-evoking events, human-computer speed mismatch, probability of computer error, probability of human error, and the level of feedback between human and computer.

The model is based on queueing theory concepts. Multi-task decision making is described in terms of events and actions. The decision maker's

task is to detect action-evoking events and to decide whether or not any particular task warrants his full attention. The model assumes the human to employ a quasi-optimal decision making strategy for scanning displays and allocating attention. However, the results of our simulation studies are probably not very sensitive to the particular strategy employed, mainly because we were most interested in gross effects.

The theoretical formulation has given us valuable insights to problems associated with the design of an adaptive decision making system. However, before we can use this formulation to evaluate potential designs, a better understanding of human decision making in multi-task situations is needed.

AN EXPERIMENTAL SITUATION

We have developed the following experimental situation with two goals in mind. First, we wanted a situation of enough generality to allow several experiments to be performed without substantial changes in the software and/or hardware. Second, we wanted to be consistent with the simulations being developed at NASA Ames Research Center [4].

Figure 1 illustrates the CRT display observed by the pilot. The display is a Hewlett Packard 1310 while the computer generating the display is a PDP-11/40 augmented by special purpose equipment for refreshing the display. The pilot's task is to fly the Boeing 707 dynamics along the map while maintaining specified altitude and air speed. His flight instruments are displayed beside the map. The pilot has aileron, elevator, and thrust controls.

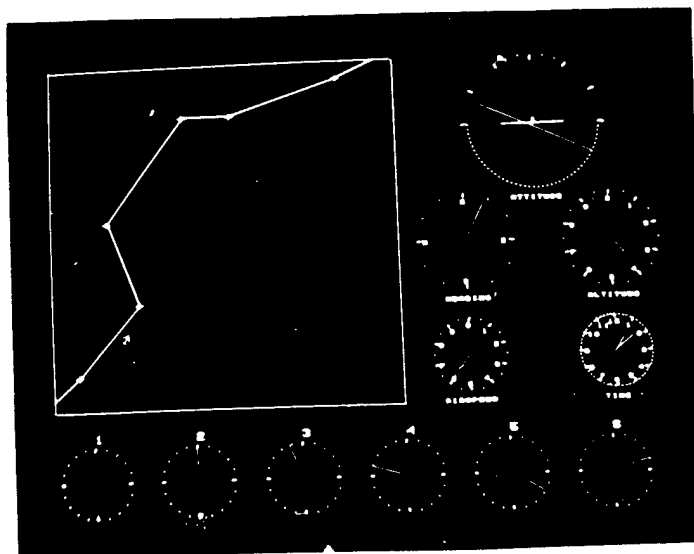


Figure 1: The Experimental Situation

The instruments below the map represent the numerous aircraft subsystems (e.g., electrical, hydraulic, etc.) which the pilot monitors for possible action-evoking events. Each of the subsystem instruments represents a linear dynamic process driven by non-zero mean white noise input. An event is defined as the removal of the input, which causes the state of the process to asymptotically approach zero (pointer down).

Upon detecting an event to which he wishes to respond, the pilot selects that subsystem via a 4 x 3 keyboard. The display shown in Figure 2 then appears. This represents the first level of a checklist-like tree associated with the subsystem of interest. He searches for a branch labeled with a zero and selects that branch with his keyboard. If the tree has more levels, the next level is then displayed. After completing the last level of the tree, the action is complete and the display shown in Figure 1 returns.

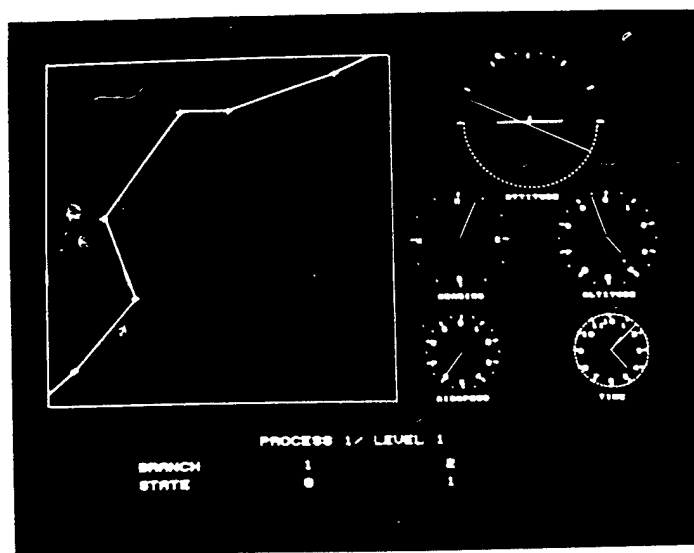


Figure 2: Display When Pilot Has Reacted to an Action-Evoking Event

Within this experimental situation, we can vary the complexity of the map, the availability of autopilot, the number of subsystems, the distribution of arrivals of events among tasks, and many other aircraft and subsystem parameters.

EXPERIMENTS PLANNED

Two sets of experiments are planned using the above experimental situation. The first set will address the question: How is pilot decision making (in terms of response times, errors, and computer usage) affected by task complexity, autopilot, and non-adaptive computer aiding?

At this point, the following scenario is envisioned. Subjects will fly 50 minutes enroute (simple map) and there will be little subsystem activity (few events). Then, they will fly 10 minutes in a landing situation (complex map) where there will be considerable subsystem activity. Experimental measurements will include:

1. Response time to subsystem events,
2. Probabilities of error,
 - a. False alarm
 - b. Missed event
 - c. Incorrect action
3. Course errors,
4. Control actions.

For the first experiment, each subject will perform three trials where the experimental variable of interest will be:

1. No autopilot,
2. Reliable autopilot,
3. Unreliable autopilot.

Reliable autopilot refers to one with a low probability of malfunction while unreliable refers to one with a high probability of malfunction. An autopilot malfunction is characterized by the aircraft maintaining its current bearing and air speed and thereby eventually deviating from the desired path. Once a malfunction occurs, a subject's task is to take over control of the aircraft, return it to the appropriate course, and re-engage the autopilot (which never fails irreparably).

The second experiment of this first set of experiments will employ the same scenario and consider pilot interaction with a decision making system designed to aid him in detecting subsystem events and acting appropriately. This system will be non-adaptive and designed on the basis of results obtained from the first experiment. The computer aid will have two modes of operation (other than "off"):

1. Detection only,
2. Detection and action.

It will also have an internal adjustment whereby its probability of successful decision making can be changed. When initialized, this probability will be high enough to be a significant aid to the pilot. However, if a malfunction occurs, the probability will decrease to the point of making

the computer aid a hinderance. The subject's task will be to detect this occurrence and reset the adjustment via a simple pushbutton. The computer aid will have two levels of reliability which will be characterized in a manner similar to the autopilot, probably using the same numerics to standardize subjects' perception of reliability.

The empirical results obtained from this first set of experiments will be combined with the theoretical formulation discussed above and employed in the design of an adaptive decision making system. Besides the queueing theory formulation, we are also considering control theory as a methodology with which to design an adaptive system. Another issue of special importance to an adaptive system is the measurement of both aircraft and pilot states. Several alternative approaches are being considered. The adaptive system that results from our investigations will be experimentally evaluated using the situation discussed in this paper.

CONCLUSIONS

It appears inevitable that aircraft will become increasingly automated. However, it also appears unlikely that the pilot will disappear from the cockpit. Instead the pilot and computer will have to cooperate in managing the aircraft. This paper has discussed an experimental situation that is being used to theoretically and experimentally determine how cooperation between these two decision makers can be enhanced.

REFERENCES

1. W. B. Rouse, "Human Interaction With an Intelligent Computer in Multi-Task Situations", Proceedings of the Eleventh Annual Conference on Manual Control, NASA Ames Research Center, May 1975, pp. 130-143.
2. W. B. Rouse, "Adaptive Allocation of Decision Making Responsibility Between Supervisor and Computer", Proceedings of the International Symposium on Monitoring Behavior and Supervisory Control, Berchtesgaden, F. R. Germany, March 1976.
3. W. B. Rouse, "Human-Computer Interaction in Multi-Task Situations", Submitted for publication.
4. E. Palmer, "Pilot's Manual for the 4D Area Navigation and Autopilot Systems in the Flight Management Project Cockpit Simulation", NASA Ames Research Center, NASA TM X-73, 100, October 1975.

AN EXPERIMENTAL SITUATION FOR STUDY OF
HUMAN DECISION MAKING IN MULTI-PROCESS MONITORING*

by William B. Rouse and Joel S. Greenstein

University of Illinois
Urbana, Illinois 61801

SUMMARY

This paper describes an experimental situation which has been developed to enable study of the human's decision strategy and performance when monitoring multiple dynamic processes. The decision maker's goal is to detect dynamic and/or statistical changes in any of the processes and take appropriate action to return the process to an acceptable mode of behavior. This situation is somewhat analogous to industrial monitoring and air traffic control. A series of planned experiments is discussed whose results are to be employed in the design of a computer-aided process monitoring system where the computer learns how to detect changes by "watching" the human perform the task.

INTRODUCTION

Quite frequently, a human must simultaneously monitor several dynamic processes. His task is to detect action-evoking events and implement the appropriate actions. As the number of processes increases and/or as the frequency of action-evoking events increases, the human becomes overloaded and overall system performance decreases.

The goal of the project discussed in this paper is the design of a computer-aided process monitoring system. Three important issues in the design of such a system are:

1. Allocation of decision making responsibility,
2. Resolution of conflicts between decision makers,
3. Human confidence in the computer system.

Theoretical studies [1,2,3] have led us to suggest that allocation of responsibility should be dynamic (i.e., situation dependent) with particular decision making tasks being assigned to the decision maker who, at the moment, has the time to devote to the task. This approach can yield significant performance benefits if conflicts between decision makers can be avoided.

* Supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAR-07-72-C-0259.

Instead of resolving conflicts, it would be preferable to avoid them. An approach to accomplishing this involves giving the computer a model of human behavior so that the computer may infer the state of the human within the overall system. To develop an understanding of human decision making in the simultaneous monitoring of several dynamic processes, we have designed an experimental situation. This situation will be employed both in the development of models of human decision making and for evaluation of potential systems for computer-aided process monitoring.

AN EXPERIMENTAL SITUATION

Figure 1 illustrates the display observed by the subject. The display is generated on a Tektronix 4010 display terminal by a time-shared DEC-System 10 computer and depicts the measured values of the outputs of nine processes over the past 100 sampling intervals. The subject's task is to monitor the processes, via the display, for the occurrence of abnormal events. The processes normally have identical second order system characteristics with zero-mean Gaussian white noise inputs of identical variance. The displayed measurements of the process outputs are corrupted by additive zero-mean Gaussian white noise sequences which normally have identical variance. Abnormal events might be represented by changes in the dynamics of a process, changes in the process input, or an increase in measurement noise variance.

After scanning the nine process histories, the subject is given an opportunity to key in the numbers of processes in which he has decided an abnormal event has occurred. He also enters his estimate of the time at which the event occurred. Upon completion of his responses he is given feedback regarding the actual states of the processes he has keyed in ("1" indicates the normal state, "0" indicates an abnormal state). The display is then erased, any abnormal processes detected by the subject are returned to the normal state, and a new display depicting the process histories advanced in time is generated as illustrated in Figure 2.

Within this experimental situation we can vary the number and type of different events, the frequency of occurrence of events, the distribution of events over processes, and the amount of time the process histories are advanced each iteration. The format in which the subject reports event occurrences and receives feedback can also be varied.

EXPERIMENTS PLANNED

Two sets of experiments are planned that will employ the above experimental situation. The first set will investigate human decision making in event detection and attention allocation. In the second set of experiments, computer-aided process monitoring will be studied.

Within the first set of experiments, one experiment is currently underway while another will soon begin. The first experiment is aimed at

understanding the human's event detection behavior. Subjects are allowed to respond to as many events as they think have occurred. Their false alarms are subtracted from their hits and then, the remainder is divided by a measure of their delay in responding (event waiting time). The resulting measure is their score which they are instructed to maximize.

To model human decision making in this task, we are employing signal detection theory and discriminant analysis. Our goal is to develop a model that will be of real time use to a computer as it attempts to assess the state of the decision maker.

The experiment that is soon to begin emphasizes event detection and attention allocation. In this experiment, subjects will only be allowed to respond to one event per iteration. Thus, they must tradeoff uncertainty and costs to reach a decision. We have developed a queueing theory model of this task and will compare it to the human's decision making performance in this experiment.

The second set of experiments will consider computer-aided process monitoring. We envision that the computer will employ the detection and attention allocation models noted above to assess the state of the decision maker. The computer will then be able to adapt to both the state of the task and the state of the human.

CONCLUSIONS

Many human tasks can be abstracted as the simultaneous monitoring of multiple dynamic processes. Computer aiding may allow the human to monitor an increased number of processes and also to monitor more effectively. The research discussed in this paper is aimed at designing and evaluating a computer-aided process monitoring system along the lines of that proposed in reference 3.

REFERENCES

1. Rouse, W. B.: Human Interaction With an Intelligent Computer in Multi-Task Situations. Proceedings of the Eleventh Annual Conference on Manual Control, NASA Ames Research Center, May 1975, pp. 130-143.
2. Rouse, W. B.: Adaptive Allocation of Decision Making Responsibility Between Supervisor and Computer. Proceedings of the International Symposium on Monitoring Behavior and Supervisory Control, Berchtesgaden, Germany, March 1976.
3. Rouse, W. B.: Human-Computer Interaction in Multi-Task Situations. Submitted for publication.

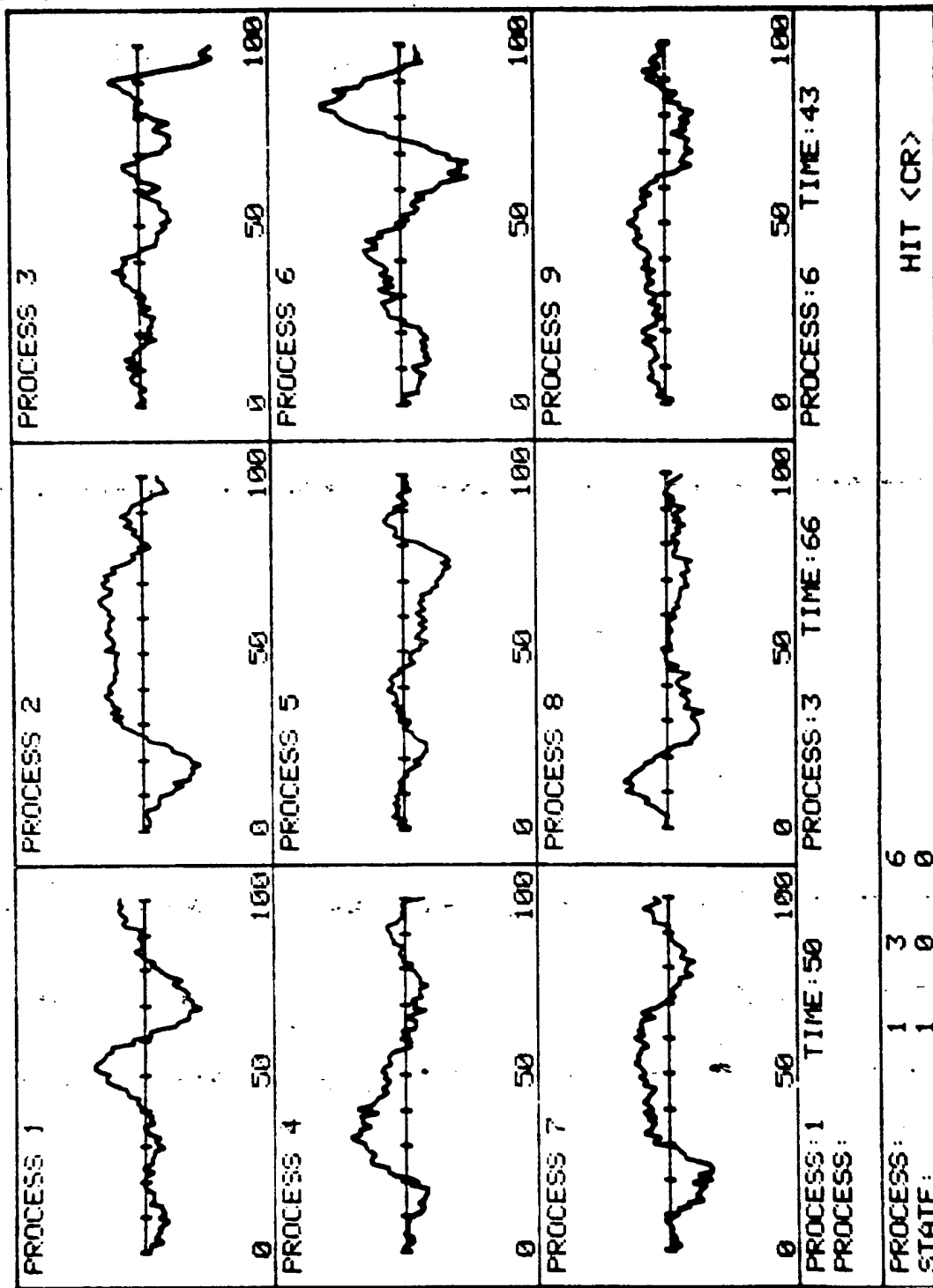


Figure 1: The Experimental Situation

THE PSYCHOPHYSICS OF RANDOM PROCESSES

Renwick E. Curry and T. Govindaraj

Man-Vehicle Laboratory
Massachusetts Institute of Technology
Cambridge, Mass. 02139

ABSTRACT

One of the major functions of the human monitor is to detect failures in systems under observation. For systems which display continuous-time variables, this is equivalent to determining when the statistical parameters of the random process have changed. This paper reports on some preliminary experiments to determine the threshold values for changes in the variance, bandwidth, and damping ratio of a second order random process. The detection time appears to be summarized by the change in the standard deviation of the velocity of the display.

INTRODUCTION

Human monitoring and supervisory control has become a research topic of increasing interest due to the advances in lower cost automation (Sheridan and Johansen, 1976). One aspect of human monitoring is that of failure detection, and some results and models have been proposed for modelling the human's ability to detect changes the mean of random process (Gai and Curry, 1976). However, there are many other parameters of a random process which are subject to change (Anyakora and Lees, 1972), examples of which are changes in variance, bandwidth, and damping ratio. In a recent review (Curry and Gai, 1976) we were unable to uncover data on the ability of the human to detect changes in these parameters. Thus we performed a set of experiments to determine the preliminary threshold values for changes in these random process parameters, i.e., to determine some of the basic psychophysical constants of random processes. These data when complete will be valuable in designing more extensive experiments, and will also be useful in modelling efforts and in obtaining performance estimates for the human in failure detection tasks.

EXPERIMENT

Overview

This experiment was conducted to determine the thresholds for various parameters of a stochastic process. The process is the output of a second order shaping filter with transfer function

$$\frac{K}{(s/\omega_n)^2 + 2(\zeta/\omega_n)s + 1}$$

and zero mean white noise input. The output was displayed by the vertical displacement of a horizontal line on an oscilloscope screen. After the steady state was reached, one of the parameters of the filter was either increased or decreased (frequency, ω , damping, ζ , or the gain, K (for noise power and hence the output variance)): the observer was expected to detect the "failure," or change. All the relevant parameters and the detection time were recorded for each trial.

Equipment

A PDP-11/10 digital computer and TR48 analog computer were used for the experiments. The uniformly distributed zero mean white noise was generated digitally and was passed through a second order (digital) shaping filter. The analog computer was then used to smooth the output for display on an oscilloscope screen. The smoothing filter pole was far away from the shaping filter natural frequency, but sufficiently low to eliminate any discrete jumps when the high frequency content of the process was increased. The subject was seated about 2½ feet in front of the oscilloscope screen. The display was scaled to such that the screen height was equal to 6σ of the process. Graticule marks provided the reference. The subject had two switches (for responding increase and decrease) to indicate his decision when the change was detected. These switches were continuously read before each update of the output of the process every 1/30 of a second.

Procedure

Sixteen graduate students participated in two sessions lasting approximately 40 minutes each. Each session consisted of three series of trials in which one of the three parameters was changed from its nominal value. Only one of the four nominal random processes was used in each session; the four nominals were obtained by the factorial combination of $T(=2\pi/\omega_n) = 1, 3$ and $\zeta = .2, .707$.

At the start of the familiarization phase, the normal or nominal mode was shown to the subject for two minutes. After this normal mode presentation, large failures were shown to familiarize the subject with the nature of the change. The change occurred randomly between 7 and 12 seconds after the start of the trial. Five practice trials were normally sufficient to allow the subject to become familiar with the changes. If the procedure was not clear at this stage the five initial trials were repeated.

Before the beginning of the experiment the subject was told of the stopping criterion, i.e., that the objective was to determine the minimum detectable change in the parameter value. He was instructed to press the switch as soon as he was certain about his decision. Correct detections were acknowledged by the printing of a single character on the teletype which was audible to the subject. When an error was made, the nature of the error was printed on the teletype and the experimenter informed the subject about the error. Three types of error were possible: if a change was 'detected' before one really occurred it was labelled a false alarm; if change was judged as a decrease while it was actually an increase (or vice versa), or if the change was not detected within the time limit, it was labelled as a missed alarm. During any trial, the failed mode continued for 30 seconds after the beginning of the failure; if a decision had not been made by this time, the motion stopped at zero and a period of 5 seconds was given to decide. If a decision was not made after this, it was considered a missed alarm. The experiment continued until he made three or more errors in the most recent six trials.

After each trial, a blanking period of three seconds was given before starting the next trial. From one trial to the next, the parameter change could be either an increase or a decrease according to the following relation

$$\left(\frac{P}{P_N}\right) = \exp \{ \ln R \times S \}$$

P_N - Nominal value of the parameter

P - Changed or failed value

R - Ratio for initial change ($R = 10$)

S - Stimulus ($S = \pm .8, \pm .6, \pm .4, \pm .2, \pm 0.16, \pm 0.12, \pm 0.04, \pm 0.03, \pm 0.02, \pm 0.01, \pm 0.008, \pm 0.006, \pm 0.004, \pm 0.002, \pm 0.0015, \pm 0.0001, \pm 0.0005$)

To avoid guessing by the subject, the decrease in magnitude, Δs , was taken for two steps, resulting in a set of four stimuli ($S_1, S_2, -S_1, -S_2$). The stimulus was chosen at random for presentation among these four until all the four were exhausted. Then the next set of four was similarly chosen and presented. The subject was told only that the magnitude of change would be decreasing in such a way that it would become progressively harder to detect its change. Initially, the changes were rather large, $S = \pm .8$ or 0.6 etc.

RESULTS

Figures 1, 2, and 3 show plots of detection time as a function of the ratio of the parameter change (on a logarithmic scale) for variance, bandwidth (period), and damping ratio. For later analysis we have plotted detection times for changes in variance and bandwidth as a function of the change in the standard deviation of the observed random process. The curves shown in Figures 4 and 5 are least squares fit to the detection time data.

Threshold values were obtained using a maximum likelihood estimate for the observed probabilities of correct responses assuming a high-threshold model for detection. The likelihood function for the observed responses is

given by :

$$L = \prod_i P(C|S_i)^{n_{c_i}} P(W|S_i)^{n_{w_i}}$$

where $P(C|S_i)$ = Probability correct, stimulus S_i

$$P(W|S_i) = 1 - P(C|S_i)$$

n_{c_i} = number correct, stimulus S_i

n_{w_i} = number wrong, stimulus S_i

The expression for probability of being correct for the high-threshold-theory model is

$$P(C|S_i) = P(C|D)P(D) + P(C|\bar{D})P(\bar{D})$$

where $P(C|D) = 1$

$$P(C|\bar{D}) = \frac{1}{2} \text{ (guessing factor)}$$

$$P(D) = \Phi \left[\frac{\Delta_i - \mu}{\sigma} \right]$$

$D \Rightarrow$ in detection state, $\bar{D} \Rightarrow$ in nondetection state

where Δ_i is the stimulus level, μ is the threshold value and σ is the standard deviation. We found that using $\Delta_i = \log S_i$ gave an appropriate Gaussian form for the stimulus values. (This may be due to the somewhat limited number of responses at each stimulus level, at the most eight).

The calculated thresholds are shown in Table 1 for the four nominal and random processes. The fit of this threshold model to the data was not particularly good because of the small number of samples available, so that these values must be considered extremely preliminary. This is due to the small number of subjects and the measurement technique which we employed to obtain a rapid determination of threshold. Our subsequent experiments will use more conventional psychophysical techniques.

DISCUSSION

The subjects had no difficulty distinguishing between increases and decreases in bandwidth and variance, i.e., all subjects would agree that an "increase" had occurred when so indicated by the experimenter. The same was not true for changes in damping ratio, which is reflected in the wide variation of detection times as shown in Figure 3. We finally arrived at a procedure of never explaining that it was an "increase" or "decrease," but just told the subjects, during the familiarization phase, that they were about to see a change of one sign or the other; we let the subjects determine whether it should be considered an "increase" or "decrease."

Examination of Figure 1 and 2 also indicates that there is an asymmetry in detection times for increase and decreases in variance and bandwidth. A

heuristic explanation for the asymmetry of detection times for changes in variance is suggested by the "exceedance limits" hypothesis. The subjects learn what values are "rarely exceeded" which correspond to (say) 2σ levels. The first time that the display exceeds this level is an indication that the variance has increased, and thus an increase of variance is indicated by the subject. On the other hand, when the variance is decreased, it takes the subject more cycles of the random process to determine that the display is not coming as close to the exceedance limit as it had before.

This asymmetry seems somewhat surprising if one considers the stimulus as a change in RMS velocity, because it can be shown that for the random processes considered here, the standard deviation of display velocity is

$$\sigma_x^* = \omega_n \sigma_x$$

If the concept of a Weber fraction holds, then we have

$$\frac{\Delta \sigma_x^*}{\sigma_x^*} = \frac{\Delta \omega_n}{\omega_n} = \frac{\Delta \sigma_x}{\sigma_x}$$

which would not imply any asymmetry in detection times. Such a model would seem appropriate since Brown (1960) has shown that the relative velocity threshold is approximately one-tenth of the absolute velocity. On the other hand, Figures 4 and 5 show that the detection times are reasonably well described by the change in absolute RMS velocity, rather than its relative value (a possibility suggested by Capt. R. Gressang).

As an alternative to the velocity effects, a normative approach to detect the changes in parameters would be to examine the residuals of the Kalman filter designed for the nominal process. If the Kalman filter is operating on the random process for which it was designed, the residuals will be white noise. Any deviation of the residuals from the white noise situation indicates a change has taken place. We have derived the equations for this increment to the residual auto-correlation function (Curry and Gai, 1976) and show in Figure 6 some typical changes to the autocorrelation function for changes in frequency and damping ratio. (In addition to the autocorrelation function shown, the Kalman filter residual has a white-noise component, i.e., an impulse at $\tau = 0$ in Figure 6.) Figure 6 shows that an increase in bandwidth yields a much larger autocorrelation component than does a decrease in bandwidth of even greater magnitude (measured on a logarithmic scale). Thus it seems as though the normative Kalman filter model for failure detection may be a reasonable basis for a descriptive model of human failure detection for these parameter changes, as well as the change in mean of a random process.

REFERENCES

1. Anyakora, S. and Lees, F.: Detection of Instrument Malfunctions by the Process Operator, *Chem Eng*, 1972.
2. Brown, R.H.: Weber Ratio for Visual Discrimination of Velocity, *Science*, 131, 1809-1810, 1960.

3. Curry, R. and Gai, E.: Detection of Random Process Failures by Human Monitors, in Sheridan and Johansen (see below), 1976.
4. Anyakora, S. and Lees, F.: Detection of Instrument Malfunctions by the Process Operator, *Chem Eng*, 1972.
5. Sheridan, T. and Johansen, G.: Monitoring Behavior and Supervisory Control, London: Plenum, 1976.

THRESHOLDS

Nominal	T^+/T^0	T^-/T^0	σ_x^+/σ_x^0	σ_x^-/σ_x^0
$T = 1, \zeta = 0.2$	1.1549	0.9326	1.1610	0.8097
$T = 1, \zeta = 0.707$	1.3072	0.9165	1.2655	0.7406
$T = 3, \zeta = 0.2$	1.2167	0.9009	1.3917	0.8710
$T = 3, \zeta = 0.707$	1.2070	0.9517	1.1299	0.7227

Table 1: Calculated Thresholds for Period and Standard Deviation Changes

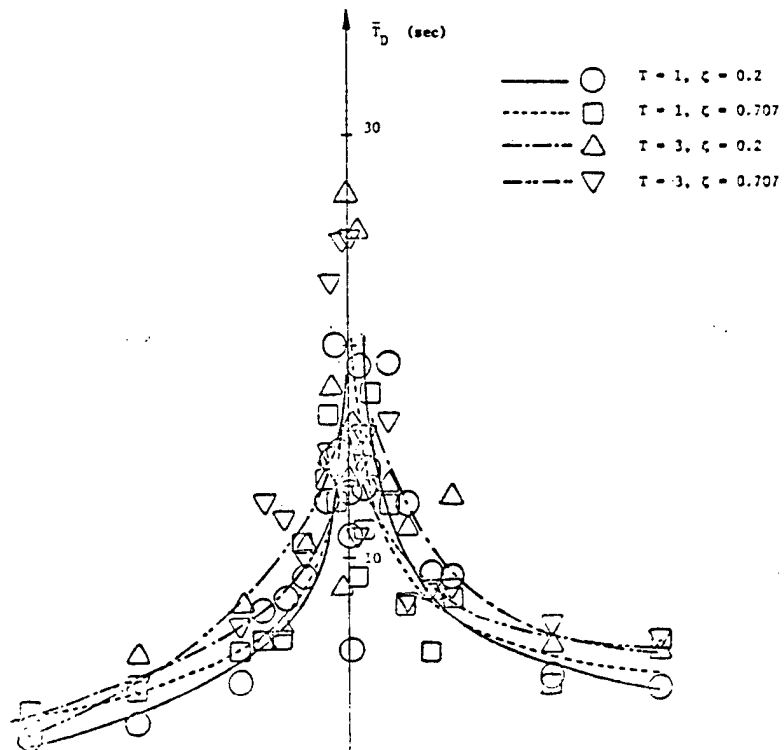


Figure 1: Detection time vs. period change

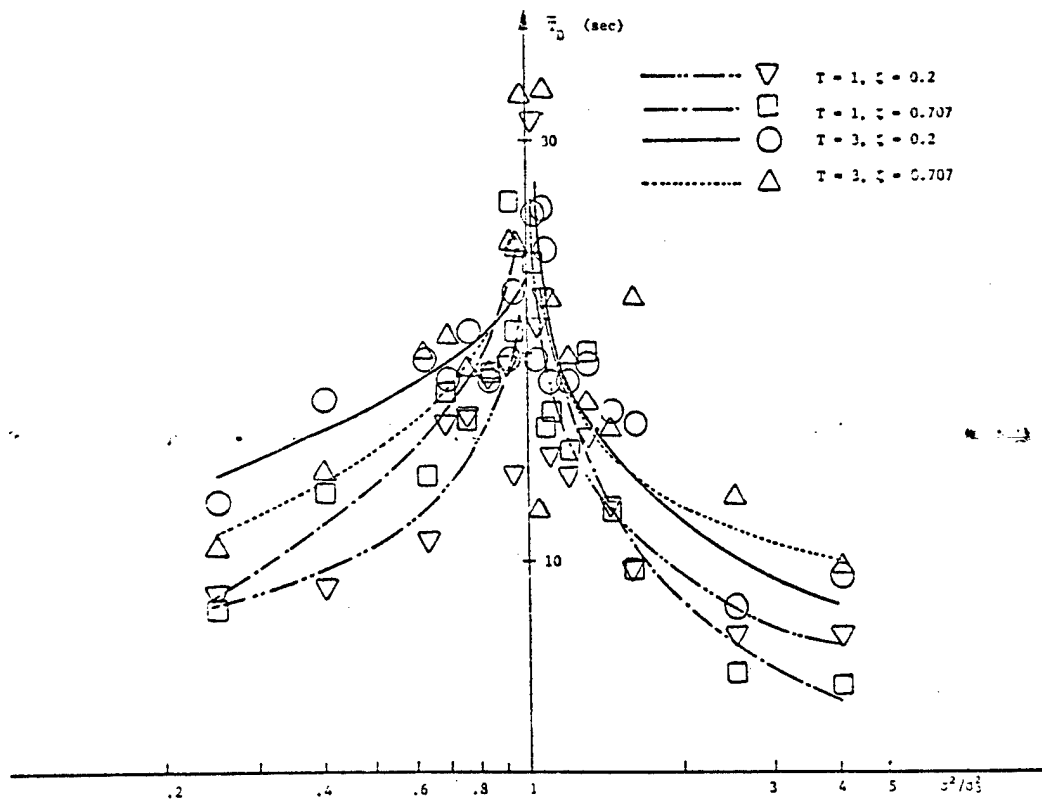


Figure 2: Detection time vs. variance change

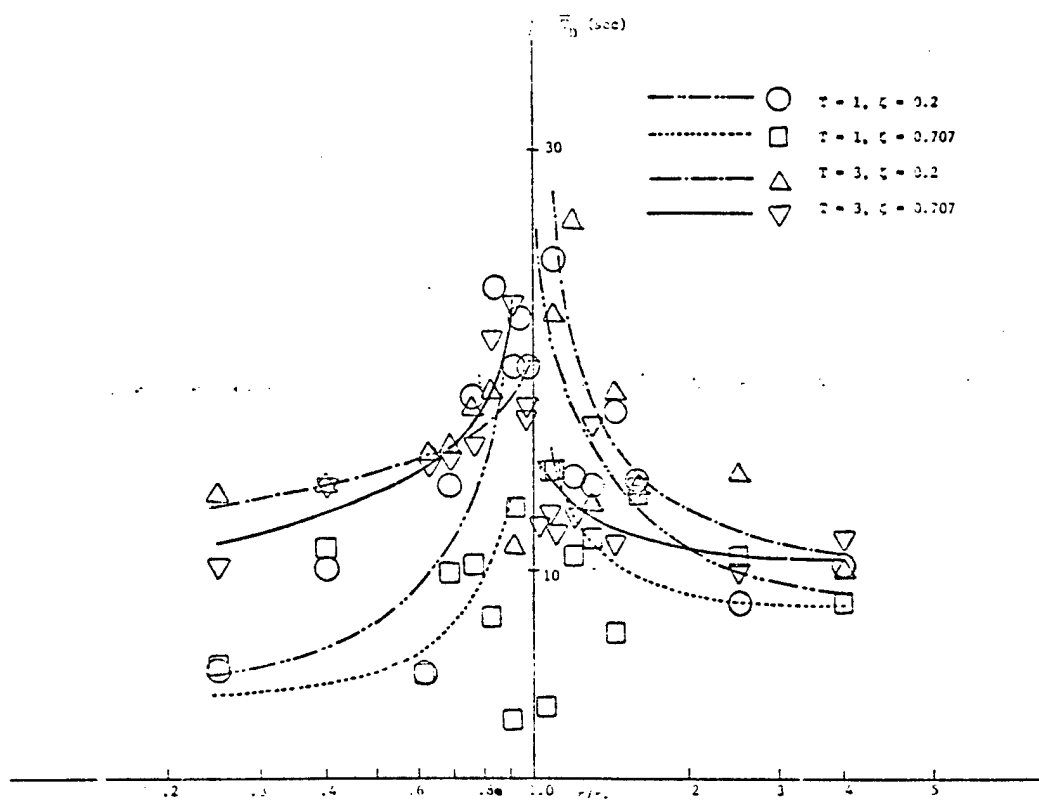


Figure 3: Detection time vs. damping ratio

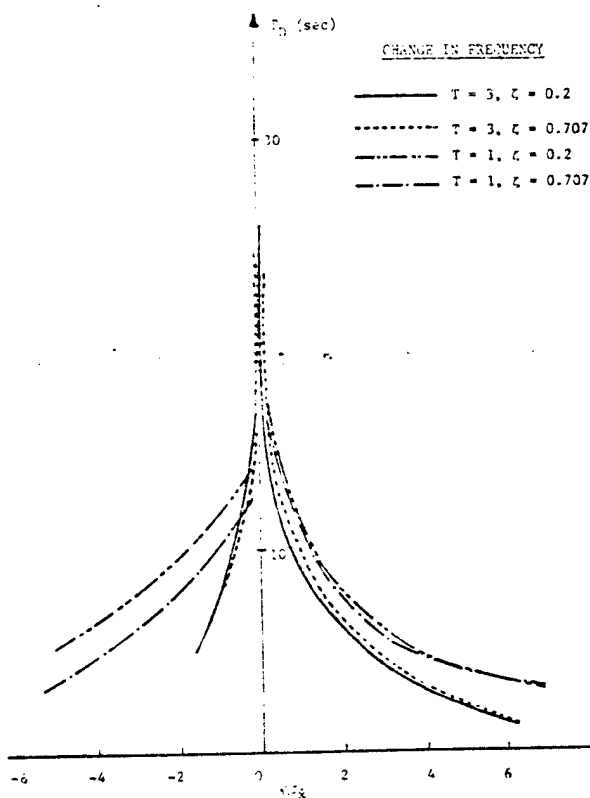


Figure 4: Least squares detection time for natural frequency changes vs. RMS velocity change

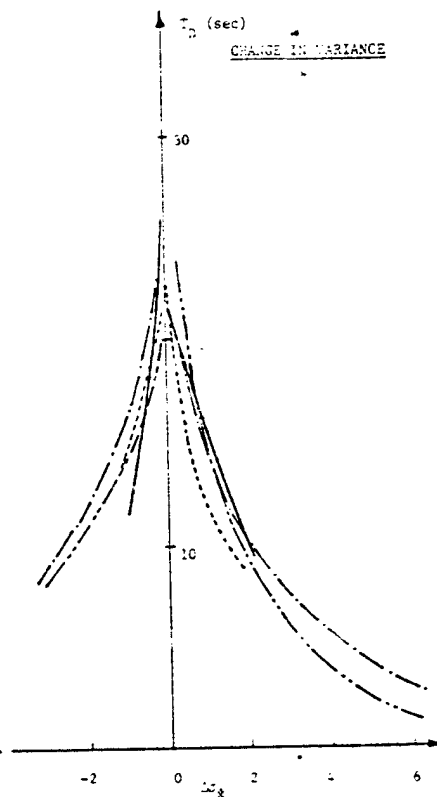


Figure 5: Least squares detection time for variance change vs. RMS velocity change

SESSION II

ATTENTION ALLOCATION AND WORKLOAD MEASUREMENT

Chairman: R. G. Mortimer

PRECEDING PAGE BLANK NOT FILMED

A NOVEL APPROACH TO THE CROSS-ADAPTIVE AUXILIARY TASK

By Arye R. Ephrath

National Research Council Associate
NASA Ames Research Center

INTRODUCTION

The technique of the adaptive task was first introduced by Birmingham (1959) and since used by various investigators in operator-performance studies (Kelley, 1962; Birmingham et al., 1962; Hudson, 1964; Kelley, 1970). Under this now-classic adaptive (or self-adjusting) scheme, the task's difficulty is modulated in real time by the operator's performance score: as the operator performs more skillfully, his task becomes more difficult. This technique found useful applications in a number of areas, such as measuring operator skill, personnel selection and training, and manual-control system design (cf. Kelley et al., 1965).

The cross-adaptive loading task technique is a modification of the self-adjusting task scheme and allows the investigator to keep the operator's performance on the primary task constant (Figure 1): the operator's primary task performance is monitored continuously and compared to a predetermined desired level; primary task performance which is better than the standard effects an increase in the difficulty not of the primary task but of a concomitant loading task, thus keeping the primary task performance constant by varying the operator's overall workload.

For the purpose of this research, however, it was considered desirable to keep the overall workload constant at a predetermined level and to study the resulting performance in a primary piloting task. Consequently, a measuring (non-loading) side task was used, and it was placed in the feedback path. This concomitant task continuously measured the operator's workload level and deviation of this measure from the preset standard, properly filtered, modulated the noise content (and hence, the difficulty) of the primary piloting task (Figure 2).

THE PRIMARY TASK

The investigation utilized a fixed-base ground simulator which was configured as the cockpit of a deHavilland CC-115 Buffalo. The primary task consisted of flying a straight, 220° localizer course while descending a 2° glide slope in zero-visibility conditions. In performing this task the subjects relied on the conventional flight instruments - airspeed, attitude, vertical speed and glideslope-deviation indicators, an altimeter, and an electronic horizontal-situation-display.

Simulated wind gusts were introduced to modulate the difficulty of this task. The gusts, both horizontal and vertical, were modelled as filtered noise of constant bandwidth and variable amplitude. This type of disturbance had been shown (Ephrath, 1975b) to affect the pilot's workload and hence, the task's difficulty.

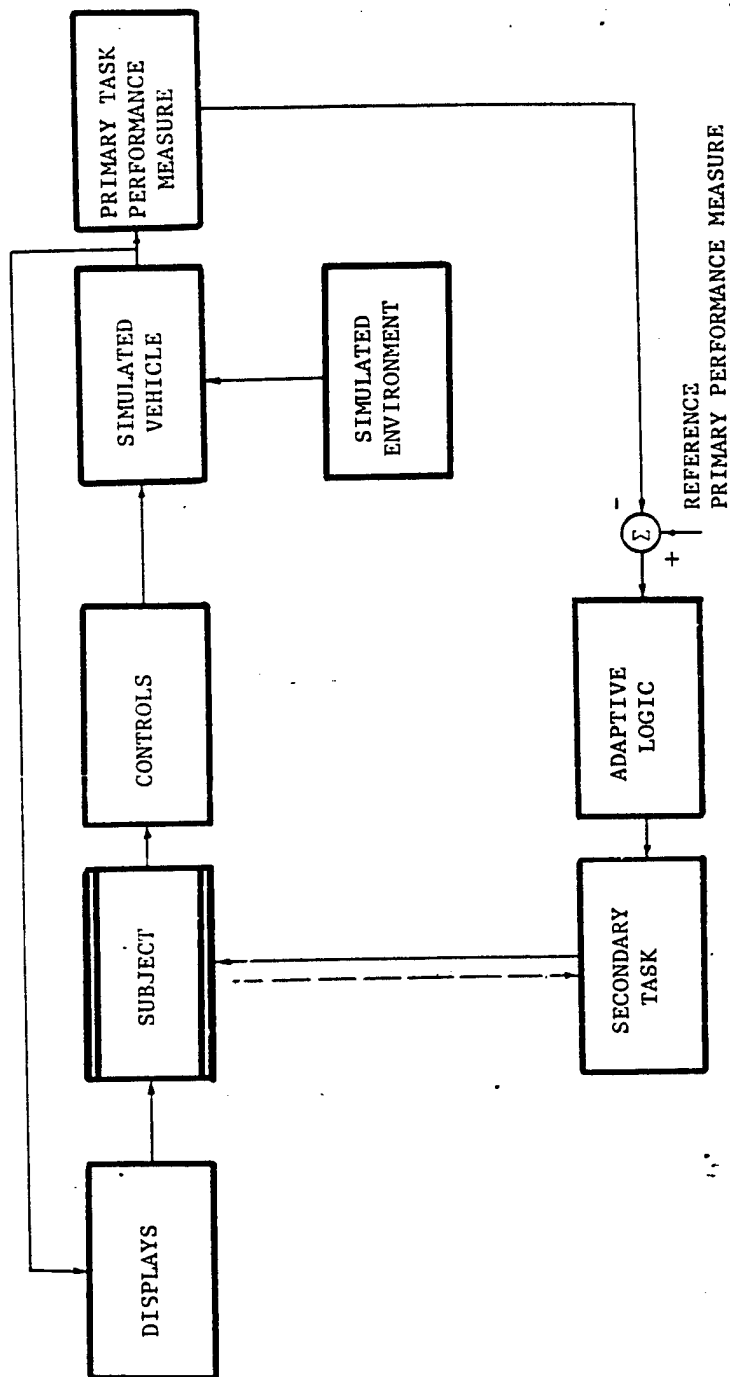


Figure 1. Conventional Cross-Adaptive Secondary Task

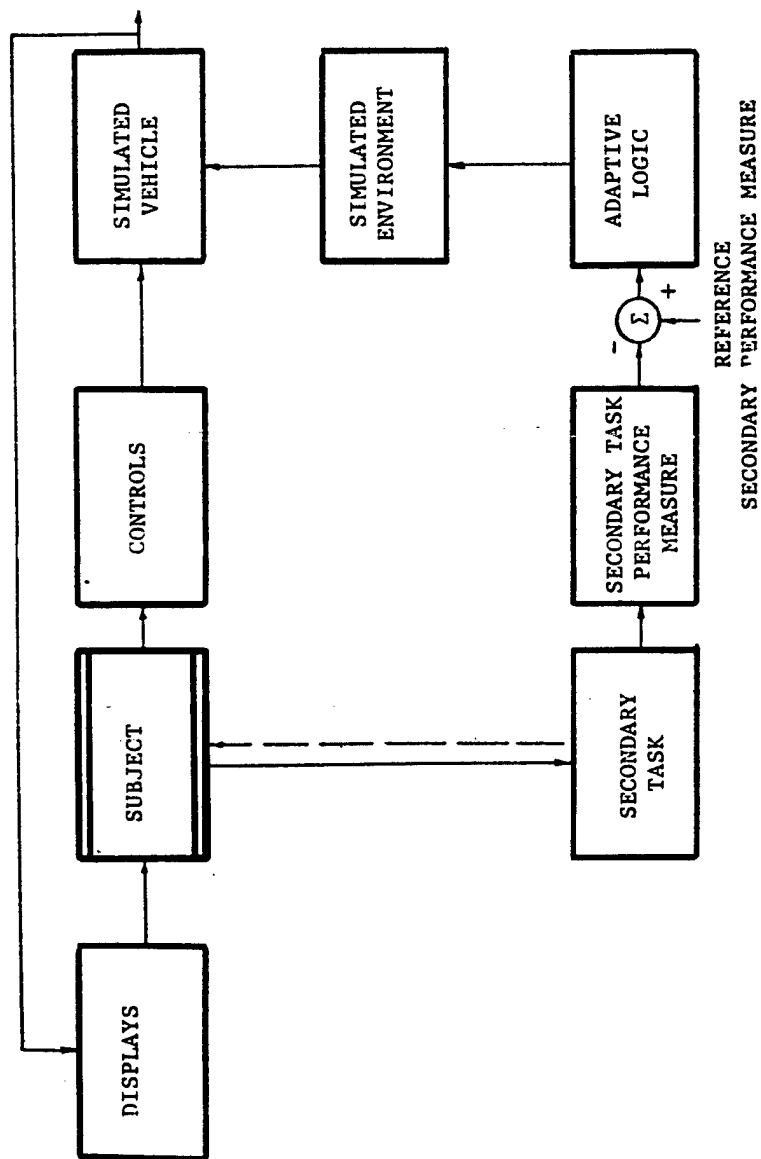


Figure 2. Novel Cross-Adaptive Secondary Task

THE AUXILIARY TASK

The auxiliary task in this study was meant to measure the subjects' instantaneous workload without introducing any significant loading of its own. It consisted of identifying, and responding to, two small red lights which were mounted above the instrument panel (see Figure 3). One or the other of these lights illuminated at random intervals, uniformly distributed between 0.5 and 5.0 seconds, and stayed on for two seconds; the subject's task was to extinguish the illuminated light by moving a toggle switch in the proper direction - up for the top light, down for the bottom light.

The lights were located 75° to the right of the center of the flight instruments. Their intensity was very low and was adjusted individually for each subject via a variable resistance to ensure that the subject would not be able to detect the lights in his peripheral vision. Consequently, the subjects could not perform both the primary and the auxiliary task simultaneously and had to switch continually between the two tasks.

The subjects were instructed to regard the piloting task as their major responsibility and to respond to the lights only if they felt that they could do so without sacrificing their piloting performance. Therefore, good auxiliary task performance was assumed to imply ease of the primary piloting task (because if the subject was faced with a difficult piloting task, he could not afford to switch regularly to the low-priority auxiliary task and, consequently, would often fail to respond to an illuminated light), and vice versa.

Each time the subject extinguished an illuminated light a "MISS_i = 0" was counted and the subject's response time RT_i was recorded. In the absence of a correct response the light stayed on for two seconds; a "MISS_i = 1" was counted and a response-time of two seconds was recorded. A pseudo-instantaneous workload index was updated after the presentation of each light, utilizing the data of the subject's response to the last two stimuli:

$$WLX_i = \frac{78.0(RT_{i-1} + RT_i) + 125.2(MISS_{i-1} + MISS_i)}{5.624}$$

which resulted in a workload-index value between 0 and 100. A measuring auxiliary task of this type has been shown (Ephrath, 1975a and 1975b) to cause minimal loading of the subject; the particular numerical coefficients chosen maximize the sensitivity of this workload measure (Spyker et al., 1971).

This workload index was compared with the desired workload level; the error signal was then integrated (to eliminate steady-state position error),

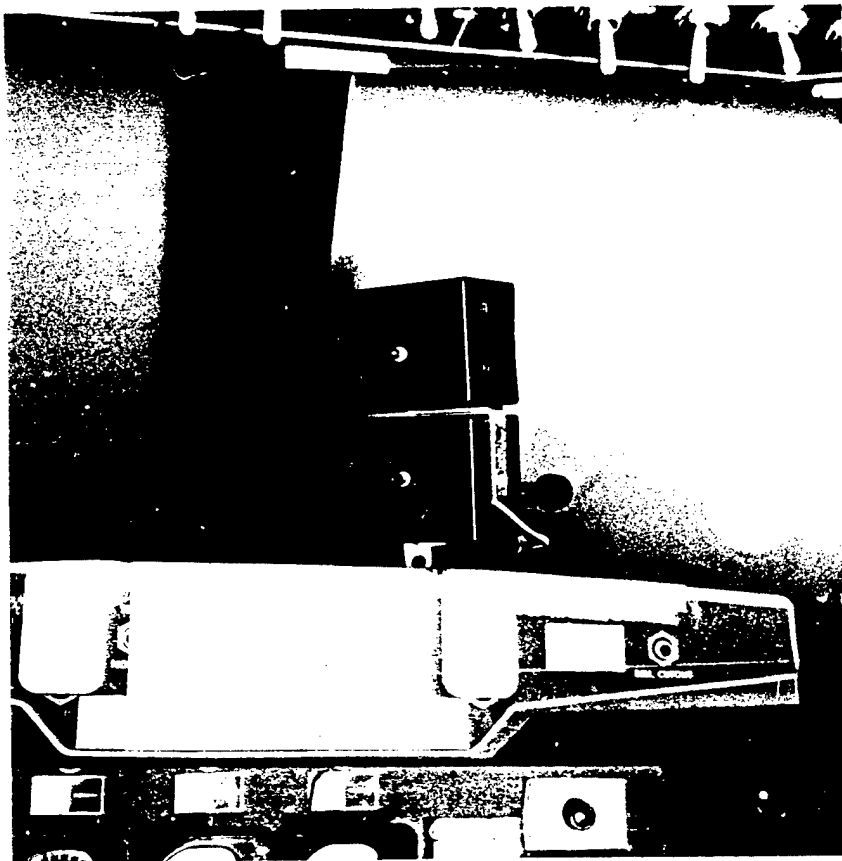


Figure 3. Side-Task Stimulus Lights

limited (to eliminate excessive gusts and to improve stability) and used to modulate the amplitude of the simulated gust disturbance (Figure 4).

RESULTS AND DISCUSSION

Each of the professional airline pilots who served as subjects flew 18 simulated approaches in the course of the formal experiment. We were aware of the fact that they might suspect a causal relationship between their performance on the light-cancelling auxiliary task and the accompanying changes in the strength of the gusts. To minimize this possibility, a value of $K_1 = 0.5$ was chosen for the integrator gain (Figure 4). This value effected a noticeable change in the gusts' strength in approximately ten seconds when the warning lights were completely ignored (implying a workload index of 100); the time lag was short enough to keep the scheme sensitive to changes in measured workload, yet apparently it was also sufficiently long to mask the correlation between gusts' strength and auxiliary-task performance: not only didn't any of our subjects detect the correlation, but neither did the colleagues who served as preliminary subjects and who might have known better or the programmers who wrote the software (without being aware of its purpose) and who had spent countless hours flying and debugging the simulator.

Workload data of a typical experimental run are shown in Figure 5. The workload index was computed from a point 46 seconds after the start of the simulated approach and is seen to oscillate around the approximate reference value WLX_{ref} . At the point indicated by the arrow a simulated guidance malfunction occurred (lateral flight director decoupled from the localizer) and the subject's workload index is seen to rise as he devoted more attention to the primary flying task. The mean workload index in this particular run was 63.6, compared to a desired (reference) workload index of 65.0. At the present time, the data of a total of 108 experimental runs is being processed for more thorough analysis.

CONCLUSIONS

This paper presents a scheme for modulating the difficulty of a primary manual-control task by means of an auxiliary task feedback. The method may find useful applications in operator workload research, in part-task simulation studies and in personnel training.

In our work, primary task difficulty was controlled by modulating the noise content of the forcing function, and our results suggest that a stable workload level may be achieved by proper selection of the subsidiary task to be used and by proper design of the adaptive logic. Other methods of controlling the primary task's difficulty - such as changing display and control gains or varying an unstable mode - may perhaps also be used; the latter of these has recently been implemented in a system evaluation study (Clement, 1976), apparently with good results.

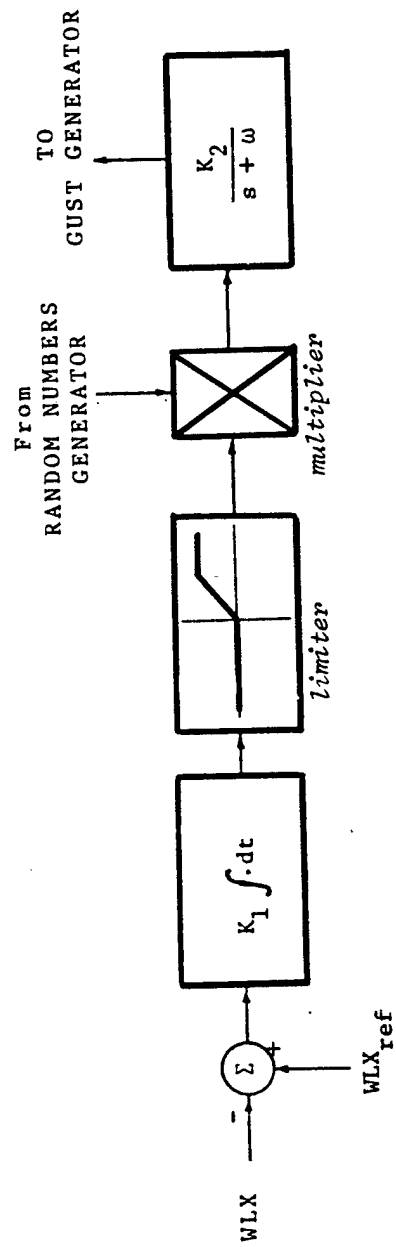


Figure 4. Adaptive Logic

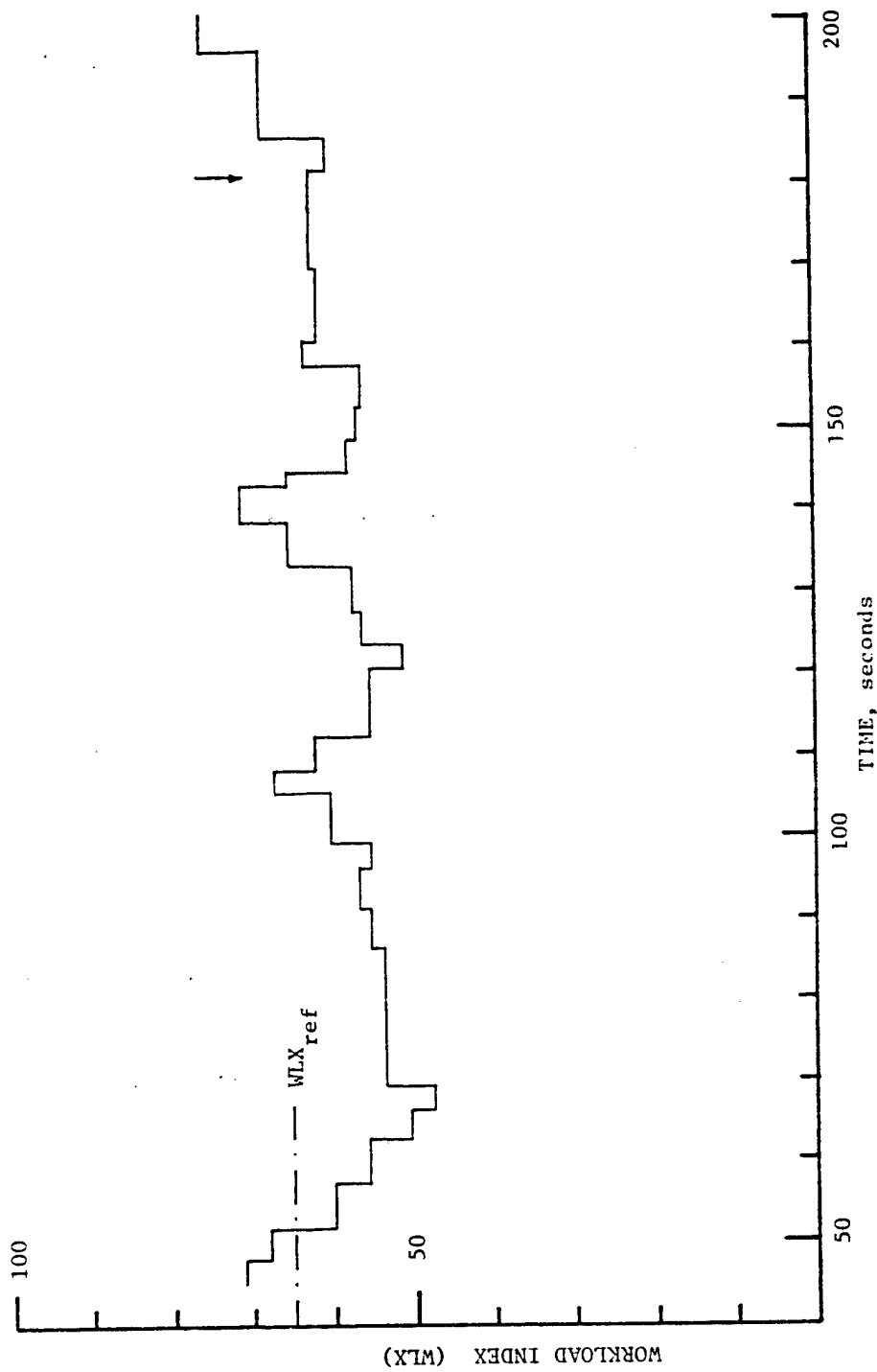


Figure 5.

Finally, this study raised, quite coincidentally, an interesting question about human perception of the correlation between control inputs and the output, its dependence on the dynamics of the plant and the effects of the associated time lags. While this question is basic to the study of human performance in manual control tasks in areas such as system identification, adaptation and fault detection, it has never been addressed explicitly, to our knowledge.

REFERENCES

1. Birmingham, H.P.: The Instantaneous Measurement of Human Operator Bandwidth. USNRL paper presented at 8th Annual Manual Control Conference, 1959.
2. Birmingham, H.P.; Chernikoff, R.; and Ziegler, P.N.: The Design and Use of Equalization Teaching Machines. IRE International Congress on Human Factors in Electronics, 1962.
3. Clement, W.F.: Investigating the Use of a Moving Map Display and a Horizontal Situation Indicator in Simulated Powered-Lift Short-Haul Operations. This conference.
4. Ephrath, A.R.: Detection of System Failures in Multi-Axes Tasks. Eleventh Annual Conference on Manual Control, 1975a.
5. Ephrath, A.R.: Pilot Performance in Zero-Visibility Precision Approach. NASA CR-137759, 1975b.
6. Hudson, E.M.: An Adaptive Tracking Simulator. IRE International Congress on Human Factors in Electronics, 1962.
7. Kelley, C.R.: Self-Adjusting Vehicle Simulators. IRE International Congress on Human Factors in Electronics, 1962.
8. Kelley, C.R.; Mitchell, M.B.; and Strudwick, P.H.: The MOL Tracking Task. Preliminary Technical Report, Dunlap and Associates, Inc., 1965.
9. Kelley, C.R.; and Kelley, E.J.: A Manual for Adaptive Techniques. Final Report, Dunlap and Associates, Inc., 1970.
10. Spyker, D.A. et al.: Development of Techniques for Measuring Pilot Workload. NASA CR-1888, 1971.

IS THERE AN OPTIMUM WORK-LOAD IN MANUAL CONTROL?

By William L. Verplank

Massachusetts Institute of Technology

SUMMARY

Experiments on a closed-circuit-television driving simulator were designed to answer the question: is there an optimum work-load that (a) sustains performance in long-term driving and (b) facilitates transitions to new tasks or emergencies? The results affirm (a) with reservations but not (b). A second experiment is planned and a new measure for work-load proposed.

BACKGROUND

Automation or semi-automation in advanced manual control systems suggests the question: Is it possible for tasks to be too easy? Most human factors engineering has been aimed at reducing work-load, but can we go too far? Is there an optimum work-load and what is the appropriate measure? The concept of arousal and the evidence of vigilance and warm-up decrements suggest the search for such a work-load measure.

AN EXPERIMENT

The task was steering a closed-circuit-television driving simulator (reference 1). The subjects were instructed to follow as closely as possible a marked circular course. Three levels of task difficulty were provided by adding different amounts of disturbance to the input of the steering servo. This task corresponds roughly to driving in wind gusts. The experiment is illustrated in Figure 1. A marker in the center of the bottom of the screen was to be lined up on the dotted center-line of the "road".

The three conditions were: A: no disturbance, B: a moderate disturbance consisting of a sum of seven sine-waves with the lower frequencies predominating, and C: a difficult disturbance consisting of a sum of the same seven sine-waves with the higher frequencies more dominant. The speed of the car was held constant throughout the experiment.

Each experimental session consisted of 14 laps (approximately 12 minutes) at one condition, followed by 2 laps of condition C, 14 laps at a second condition followed by 2 laps of condition C, and 14 laps of the third condition followed by 2 laps of C. The subjects were familiarized with each of the three conditions prior to the experiment, and during each session were warned over an intercom, approximately 5 seconds before each change in condition, what

This is an informal paper presented at:
12th Annual Conference on Manual Control
University of Illinois, 25-27 May, 1976.

the new condition was going to be. The change in conditions always occurred at the same point on the track. The whole session lasted approximately 45 minutes. Data were recorded for three subjects over two or three sessions each, with the conditions presented in a different order each time. (See Figures 2-a through g for the orders.)

Each session provided data on the six areas of interest: prolonged performance under three conditions, and adaptation to condition C after these three different conditions.

Instrumentation

Both speed-control for the vehicle and generation of the disturbance were provided by a PDP-8 computer (10 samples/second). The lateral position of the car on the track was recorded on a chart recorder. It was from this record of position error that performance scores were calculated.

Data

The absolute value of the position error was read from the chart by eye at 32 equally spaced intervals for each two-revolution block of data. Thus, if the standard deviation of error in individual measurements from the chart were 20% (with no average bias) then the two-lap average absolute position error is good to approximately 3%. These two-lap average absolute position errors are plotted versus time in Figures 2-a through 2-g. The units are centimeters, measured at the display (T.V.). Each lap took 51.2 seconds, so the two-lap averages represent 102.4 seconds of data.

RESULTS

Hypothesis I: Long-term Performance

To test the hypothesis that performance will deteriorate with time, the difference in averages at the beginning and end of each condition were calculated. Each condition lasted for seven blocks. Figure 3 shows the change in average error from the first 3 blocks to the last 3 blocks for each condition (i.e. the first five minutes versus the last five minutes in a 12 minute run). Lines connect the data points produced in the same experimental session.

There is no significant decrement in performance for conditions A or B, but there is for condition C ($t=5.9$, $df=6$, $p<.005$). These data only support half of hypothesis I, that moderate is better than difficult. Hypothesis I suggests that, as the task gets easier (condition A), there will be a tendency for a vigilance-type decrement. Either condition A was not monotonous enough, or the run (12 minutes) not long enough, or for this particular driving simulation there is no such thing as an optimum task difficulty.

There does appear to be some difference however between conditions A and

B, as they relate to long-term performance. As shown in Figure 3 there is more spread in the changes for A than for B. That is, under condition B, average performance did not change over the run, but in condition A, it did. On some sessions it deteriorated over the run, and on some, it improved. This suggested that if stability of performance is important, condition B may well be superior to condition A.

This was examined by computing the standard deviation of the two-lap averages for each 12 minute run. The normalized standard deviation is shown in Figure 4. Here is one measure that shows the moderate condition B as better than the easy condition A (paired $t=4.8$, $df=6$, $p<.005$). There is less variability in performance over the 12 minute run. However, the difference between conditions B and C disappears.

Thus, given the present data, it cannot be shown that a task of moderate difficulty is superior in long-term performance to both easier and more difficult tasks. It is, however, superior in certain ways. The moderate task B did not show a performance decrement over time, whereas the difficult task C did. The moderate task B showed less relative variability than did the easier task A.

Hypothesis II: Adaptation

Figure 5 shows error scores on condition C immediately following the extended run of either condition A, B or C. The two-lap error score is broken down into $\frac{1}{2}$ -lap scores. Figure 5-a shows the score for the first $\frac{1}{2}$ -lap (12.8 seconds). Figure 5-b averages the score for the first $\frac{1}{2}$ -lap with that for the second (average for first 25.6 seconds). Similarly 5-c and 5-d are averages of all the performance up to the specified points.

Hypothesis II would suggest that initial performance on the new task would differ with different preparatory conditions, and that upon "adaptation" to the same new task, performance would equalize. Just the opposite appears to be the case. The preparatory conditions cannot be distinguished on the first $\frac{1}{2}$ -lap score - only on the entire two-lap averages (5-d).

It is difficult to draw conclusions from the data because of the large inter-subject variability. In the two-lap averages, the only significant difference is for subjects JK and JM, between conditions B and C (paired within sessions) ($t=5.16$, $df=4$, $p<.01$). An alternative to Hypothesis II is proposed to explain this difference, and depends on a detail of the experimental procedure.

The experiment compared performance in adapting to condition C following an extended run of either A, B or C. If the "adaptation block" was followed by the extended run of C, this "adaptation" block was counted as the first block in that extended run. For "adaptation" following condition C, the block examined was simply the last block in the extended run. This may explain the similarity between Figures 5-d and 3, at least as far as condition C is concerned. Figure 3 shows the difference between the beginning and end of the extended run. For condition C, it shows a significant decrement in performance.

Similarly, Figure 5-d shows the averages at the beginning of a run (C following A or B) and at the end of a run (C following C).

There are two reasons then that the data do not support Hypothesis II. First, if it is truly a process of adaptation, the difference in performance should be greatest at the beginning of the new condition. The only significant difference is in the direction predicted by Hypothesis II (C worse than B), but it is not at the beginning, showing up only in the two-lap average. Second, this difference can be accounted for by an alternative to Hypothesis II. Another possibility is that a different measure of the "preconditioning" work-load could explain the differences in adaptation.

SUMMARY

Two hypotheses were proposed relating the concept of arousal and the evidence of vigilance and warm-up decrements to automobile driving. There should be an optimum work-load that sustains performance on long tasks, and facilitates transition to new tasks.

A simulation of automobile driving under three levels of difficulty showed a moderate disturbance caused less decrement than a difficult disturbance and less variability than no disturbance. No condition was superior at facilitating transitions.

The present evidence is not very strong support for the concept of an optimum work-load. Performance decrements in the "under-load" condition probably cannot be expected over such short runs. A second experiment with longer runs is proposed. The question of what constitutes optimality is also open.

Finally, the measure of work-load ought to include a kind of "self-loading" or "operator-induced work-load", especially in the case where there is no external disturbance and the operator is "tracking his own noise". The second experiment and a new measure of work-load will be reported in a forthcoming Ph.D. thesis [2].

REFERENCES

1. Rudy, R.P. and Willis, G. Monitoring and Control of an Automobile Simulator Using a Real-time Digital Computer. M.I.T. Engineering Projects Laboratory Report, DSR 79728-2, 1969.
2. Verplank, W.L. Sources and Effects of Uncertainty in Sustained Manual Control. Ph.D. Thesis, Mechanical Engineering Department, M.I.T. (in progress).

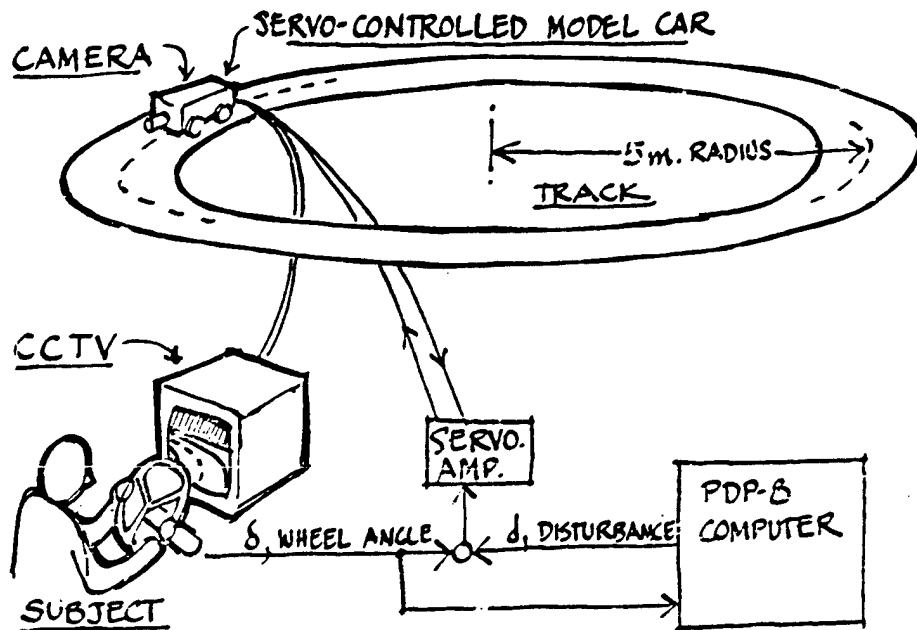


Figure 1. The M.I.T. C.C.T.V. Driving Simulator.

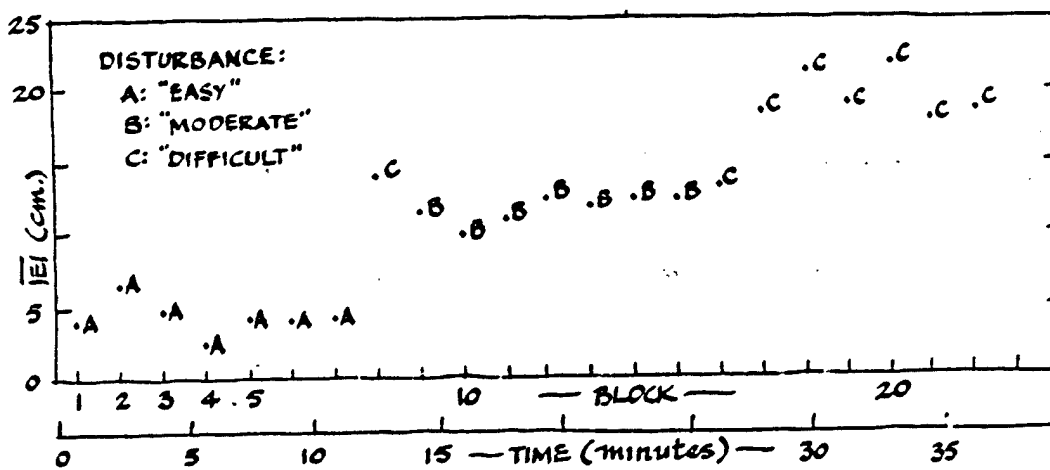
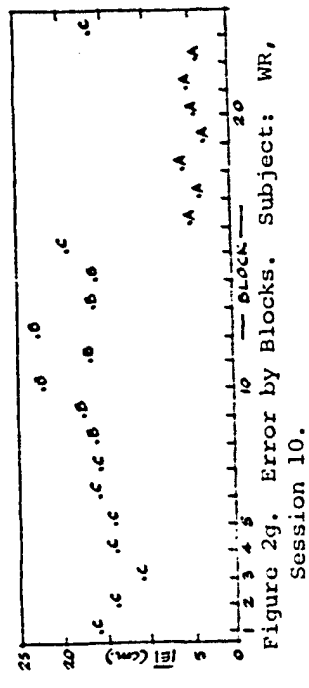
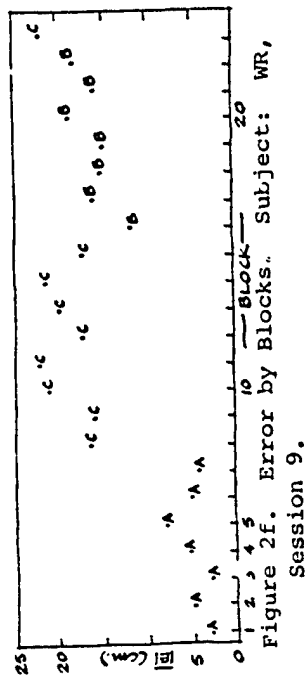
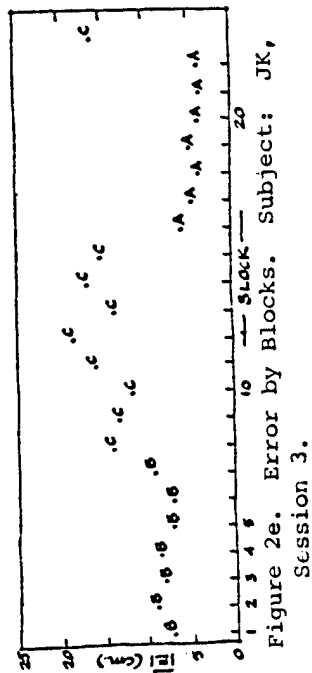
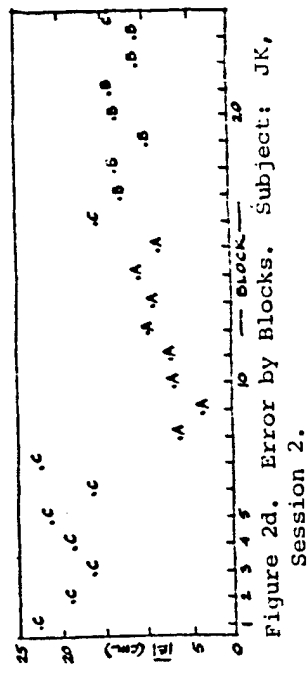
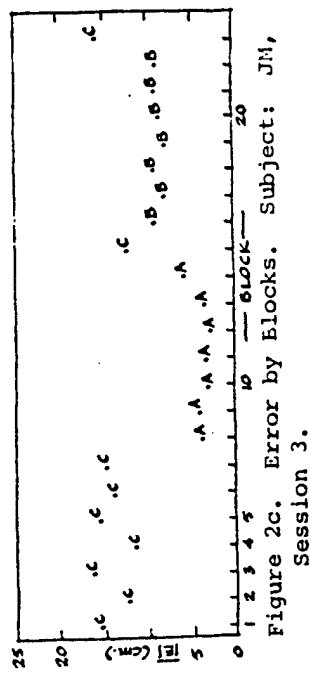
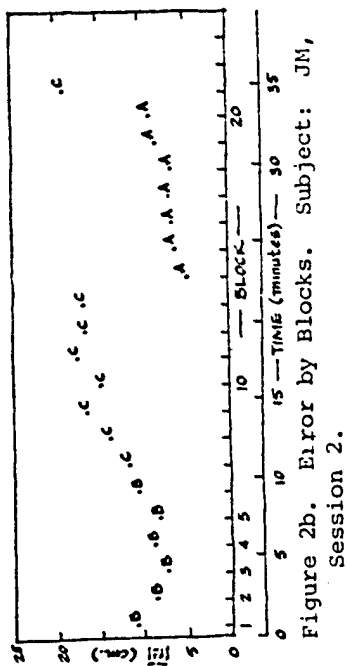


Figure 2a. Average Absolute Error by Two-lap Blocks.
Subject: JM, Session 1.



ORIGINAL PAGE IS
OF POOR QUALITY

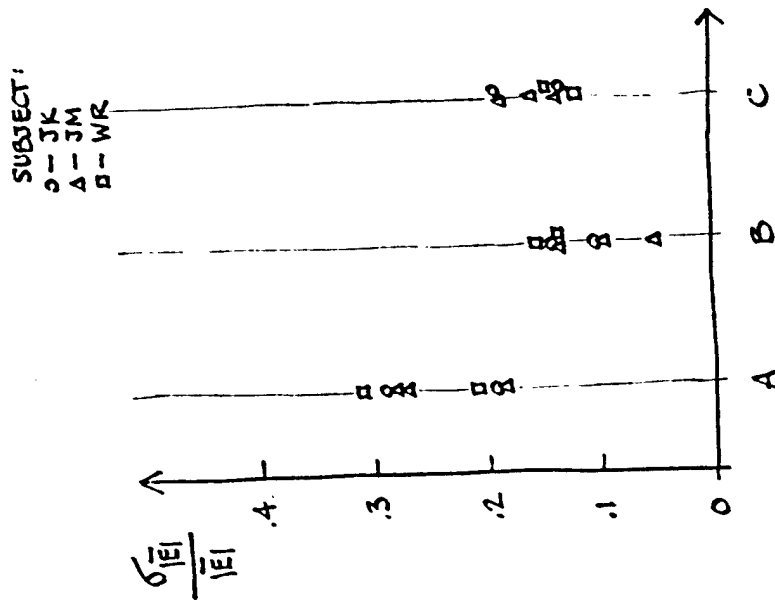


Figure 4. Relative Standard Deviation of (2-minute) Average Absolute Errors

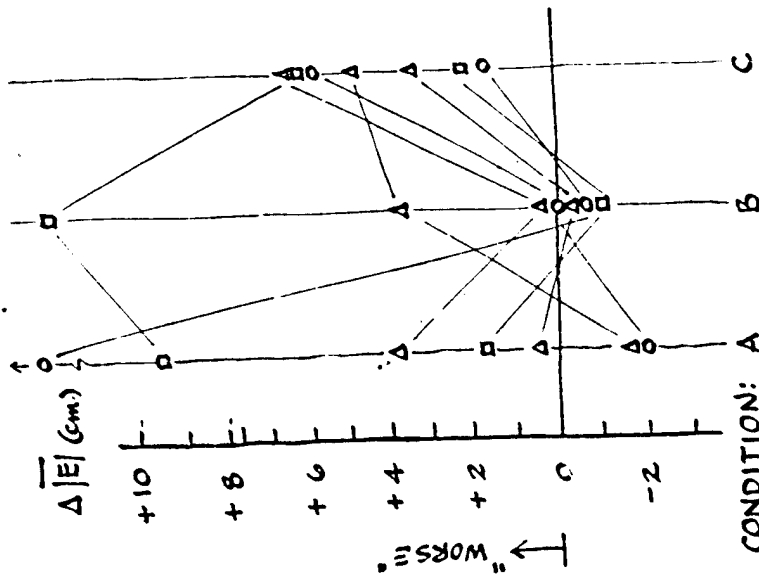


Figure 3. Change in Average Absolute Error From First- to Second-Half of 12 minute Run.

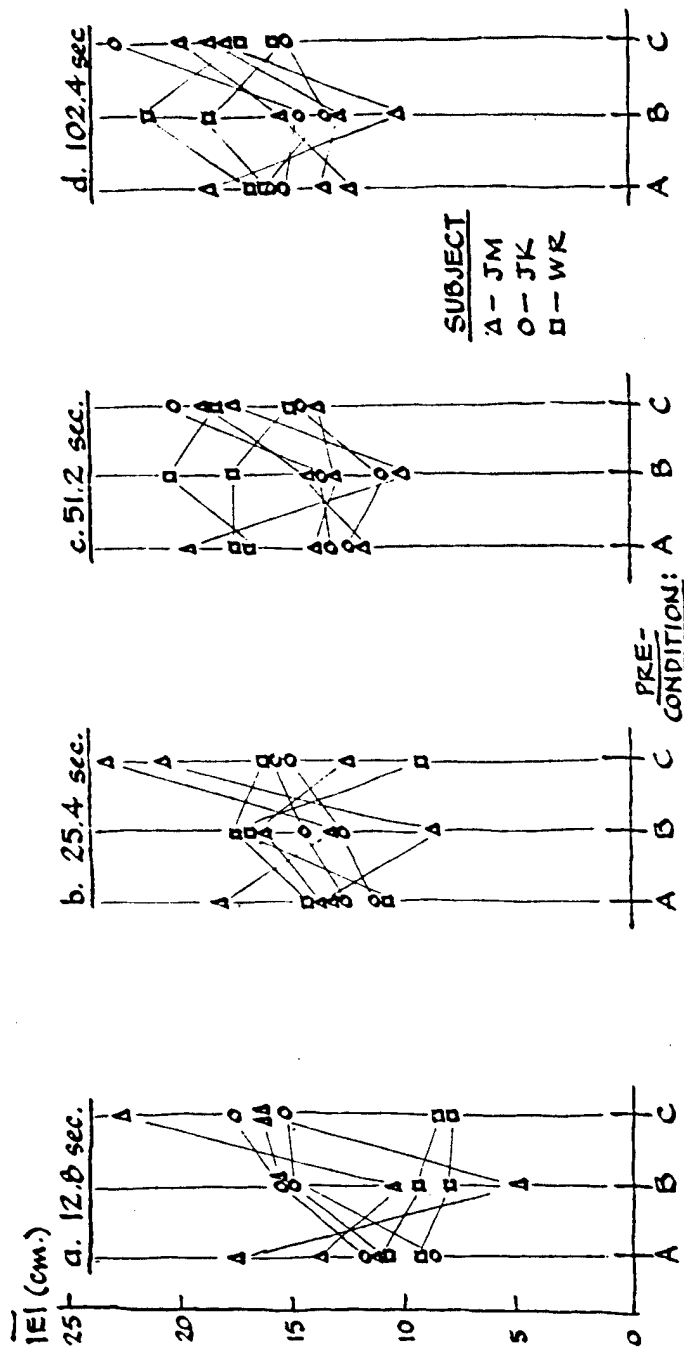


Figure 5. Average Absolute Error on Condition C Following A, B or C.

TASK INTERFERENCE IN MULTI-AXIS AIRCRAFT STABILIZATION

By Edward D. Onstott

Controls Research

Northrop Corporation
Aircraft Division
Hawthorne, California

ABSTRACT

A time domain attention allocating multi-axis pilot model has been used to examine task interference effects in a two-axis attitude stabilization task in turbulence. Configurations from a matrix of uncoupled lateral and longitudinal F-5 dynamics were analyzed to determine the influence of each axis on the control performance of the other. The analysis showed that an inappropriate choice of longitudinal dynamics would deteriorate the roll stabilization with no change in the lateral dynamics or pilot model. Furthermore, the influence of roll versus pitch angle display scalings were analyzed by hypothesizing that the task urgencies used to allocate attention in the pilot model should be weighted by the scaling factors. A fixed base simulation verified the model, the resonant task interference, and the scaling hypothesis; further validation through manned simulation was performed using a complete nonlinear YF-17 aircraft model and the Northrop Large Amplitude Simulator.

INTRODUCTION

Failure to match the dynamics of the pilot and the aircraft in an optimum way may result in not achieving the best performance of both. The result of such dynamic mismatch is greater pilot workload, less accurate weapon delivery, riskier approach and landing, and increased development cost deriving from overdesign in an attempt to achieve better pilot acceptance. Furthermore, the lack of precise methods for analytically evaluating the dynamic pilot-aircraft interface in the preliminary design phase may lead to extensive and unnecessary flight simulation. The analytic ability to recognize and correct significant control shortcomings as well as to identify potential improvements is important at a time when basic design concepts can be freely adapted to new requirements.

This analytical requirement has been partially met for decades by the use of mathematical models of human pilot dynamics. These techniques have been successfully employed in the development of simple low authority control systems where greatly simplified representations of the aircraft and the pilot could be tolerated, References 1-3. However, the advent of advanced tactical aircraft with the availability of high authority, high order control concepts and the associated demands for superior piloted performance requires pilot model aircraft analysis technology that is considerably more general.

In order to meet the current requirements of flying qualities prediction and evaluation, Northrop has evolved a flying qualities analysis procedure including complete generality of the nonlinear aerodynamics and control system. Digital simulation has been adopted as an over-all context for the dynamic representation and computation. With the availability of point-by-point generation of the flight history, pilot models can be developed that take into account not only the statistical features exploited by describing function or optimal control methods, but specific dynamic and decision processes as well. Northrop has validated this digital simulation approach to time domain decision models which generate the characteristics of multi-axis piloted flight, References 4 and 5. In addition to the increased system generality, an important consequence has been the discovery of multi-axis task interference effects that have never before been identified or analytically computed. The objective of this paper is to present these results.

ATTENTION ALLOCATION AND THE MULTI-AXIS HUMAN CONTROLLER

A great deal is known about the dynamics of the human pilot performing continuous linear single axis tasks. Much work has gone into developing models that match the amplitude and phase characteristics of the pilot's output at the controller, and many aspects of the internal structure of the human have been analyzed. These "ultra-precise" models are of use in solving many human factors problems about the interface between pilot and controller, but for the basic objective of determining the total system dynamics, it is usually sufficient to employ simple models that consist of gain K_p , lead T_L , time delay τ , and possible lag T_I :

$$Y_p = K_p \frac{(T_L s + 1)}{(T_I s + 1)} e^{-\tau s} \quad (1)$$

The more exact pilot models can certainly be employed, but for most purposes the above simple model gives good statistical results.

There have been three main approaches previously taken in attempts to extend single-axis model theory to multi-axis tasks. All of these recognize that the human must operate as a time shared device when faced with difficult control tasks on several independent axes. This shifting attention allocation degrades the performance of each axis from what the pilot would achieve in continuous control. As might be expected, these three approaches are 1) decrease the model gain from the optimum for continuous control, 2) increase the time delay to account for the periods of inattention, and 3) inject filtered noise to imitate the spectral content of the shifting pilot control.

The insufficiency with these approaches is this: the human pilot is quite discriminating about when he will abandon the control of one task to take over the control of another. This leads to a pilot sampling criterion that is functionally dependent on the total system variables. In no way can this be regarded as a purely random, or a regular sampling. Thus a multi-axis pilot model must contain an algorithm that determines when attention shifting takes place, and the model must be computed in a way that preserves this information. Recently, Northrop developed a multi-axis pilot model which does just that, the urgency function model.

By using the method of digital simulation, the exact functional criterion, by which a pilot decides his control, can be directly computed without the gross distortions of linearization. The development of the form of these urgency criteria has now advanced to the point where they can be determined from 1) the system dynamics, 2) the task, and 3) the appropriate human factor information about the pilot.

Let x_i be the state variables of one axis, x , of a two axis task, and let the other axis, y , be represented by y_i . Then the attention allocation criterion for the x axis is satisfied identically with the inequality

$$U_x(x_i) \geq U_y(y_i) \quad (2)$$

where U_x and U_y are the urgency functions of the x and the y tasks. These functions are always nonlinear in the state variables, but fall into several precise classes. Some of these classes have been well explored, and a tabulation of the urgency functions for attitude stabilization is included below.

The multi-axis urgency function model thus consists of simple linear dynamics, equation (1), along with the control criterion of (2). Whichever axis has the larger urgency function gets the corrective control attention. The adjustment of the linear coefficients can usually be obtained by an easy search starting with the optimum single axis coefficients. Almost always, the optimum multi-axis coefficients differ significantly from these values, especially in multiloop control on the two axes.

TWO-AXIS ATTITUDE STABILIZATION IN TURBULENCE

It is natural to inquire about applications of the Northrop urgency model to representative tactical fighter dynamics. Since attitude stabilization in the presence of low-level turbulence is an almost ever-present requirement of any Class IV mission, this task was selected for both analytical and experimental study using degraded F-5 aircraft models. The primary objective was to verify the time-sharing of the human pilot, the appropriateness of the urgency function attention allocation algorithm, and the statistical accuracy of the predictions. In order to control the number of parameters involved in optimizing the model, display and controller effects were minimized in the flight simulation. This involved using a large CRT display so that display motions were amplified enough to eliminate visual threshold effects. Recently a general study of this problem, including visual threshold effects, has been completed using the general YF-17 aircraft model. The results of this problem are also discussed.

The simulation display consisted of an illuminated dot against a dimly illuminated grid. Vertical displacement of the dot represented pitch angle, and lateral displacement bank angle as shown in Figure 1.

The actual scaling of the dot displacement was achieved experimentally by going to the most sensitive scope setting which would accommodate all dot excursions. These scalings, μ and α , had units of degrees per centimeter of dot deflection, and provision was made to simulate at various values of μ to α ratios. As this μ/α ratio

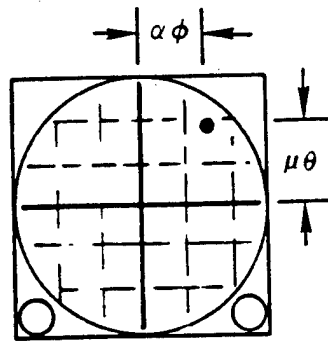


FIGURE 1. TWO-AXIS FLIGHT SIMULATION DISPLAY

increased from zero (continuous lateral tracking only), the pilot was forced to pay greater and greater attention to the longitudinal task. One of the most interesting hypotheses tested was the manner in which this took place. In all flight simulation tests the pilot was instructed to minimize the rms distance of the dot from the center of the display. This is called the radial error, denoted by $r(\phi, \theta)$, and is given by

$$r(\phi, \theta) = \sqrt{(\alpha\phi)^2 + (\mu\theta)^2} \quad (3)$$

In this way the μ/α ratio greatly influences the objective urgency, and the hypothesis tested is that the objective urgency as determined by the μ/α ratio scales exactly as the ratio of the urgency functions of the multi-axis pilot model U_P and U_R ; that is,

$$\mu/\alpha = U_P/U_R \quad (4)$$

Limitations in display resolution allowed only the range of values shown in Figure 2.

μ	α	Task
0	1	Continuous Lateral Only
8	1	Two axis, $\mu/\alpha = 8$
16	1	Two axis, $\mu/\alpha = 16$
1	0	Continuous Longitudinal Only

FIGURE 2. URGENCY RATIOS TESTED

It should also be remarked that the aircraft dynamics and data reduction of the flight simulation were digitally generated. A frame time of 0.05 second was used, and the turbulence was generated using a digital random noise source with a 0.15 kHz bandwidth, thus assuring independence of the two gust time histories produced by the digital filters. The equations of motion and all dynamic computations were programmed exactly as in the analytical digital simulation model described below. It should be noted that the method is in no way restricted to linearized equations of motion. In the YF-17 study general nonlinear six-degree-of-freedom equations were used with no change in the pilot model or its use. Three degrees of freedom were mechanized laterally, while the u equation was not required longitudinally, thus giving five degrees of freedom in all. The equations of motion are shown in Figure 3.

$$\dot{w} = \frac{1}{1 - z_w} (u_o + z_q) q + z_{\delta e} \delta e + z_w (w - \delta_{gw}) \quad (5)$$

$$\dot{q} = M_w \dot{w} + M_q q + M_{\delta e} \delta e + M_w (w - \delta_{gw}) \quad (6)$$

$$\dot{\theta} = q \quad (7)$$

$$\dot{r} = N'_r r + N'_p p + N'_{\delta a} \delta a + N'_{\delta r} \delta r + N'_{\beta} (\beta + \delta_{g\beta}) \quad (8)$$

$$\dot{\beta} = y_p^* p - r(1 - y_r^*) + \frac{g\phi}{u_o} + y_{\delta a}^* \delta a + y_{\delta r}^* \delta r - y_v^* (\beta + \delta_{g\beta}) \quad (9)$$

$$\dot{p} = L'_p p + L'_r r + L'_{\delta a} \delta a + L'_{\delta r} \delta r + L'_{\beta} (\beta + \delta_{g\beta}) \quad (10)$$

$$\dot{\phi} = p \quad (11)$$

FIGURE 3. EQUATIONS OF MOTION

The augmentation of each configuration was included in the stability derivatives, since washout is not required for a zero mean small perturbation attitude stabilization task.

The task proved difficult to fly, and about 10 hours were required for asymptotic training. Data were collected for test periods of 30 seconds, and simulation sessions were held to two hours maximum with frequent rest periods. In addition to the digital data reduction of tracking statistics, strip chart recordings were collected for each configuration and each μ/α ratio tested. In order to provide a variety of F-5 configurations, three longitudinal and two lateral sets of dynamics were selected and combined in a matrix of six configurations, designated as shown in Figure 4.

The turbulence simulation used Dryden spectra obtained by filtering Gaussian white noise as described in Mil-F-8785B, where an air speed of 718 fps and an altitude of 1750 feet were used. In order to diminish the effects of drift, a 40-db/decade high-pass prefilter was used on the random number noise source, with the break point set at 0.3 radian. The hand controller was mounted on the side of a chair and had light breakout and gradient forces; both the sensitivity and the polarity were selected by the subjects.

Longitudinal	Lateral	
	A	B
	1	1A 1B
	2	2A 2B
	3	3A 3B

FIGURE 4. F-5 CONFIGURATION DESIGNATION

Analytically, this problem was studied by means of digital simulation. This system model includes the six-degree-of-freedom linear or nonlinear airframe equations of motion, turbulence and command tracking generation, linear pilot model gain, lead, lag and delay, urgency function switching algorithms, pilot remnant, pilot inadvertent crossfeed, rms statistics, histograms of control episodes, and urgency function delay.

Since the turbulence provided a large disturbance to the system, it was not necessary to model remnant and inadvertent crossfeed. The commands to the system are simply $\phi = 0$ and $\theta = 0$.

In order to validate the model as a predictive method, the prediction algorithm should be understood. The various constants and functional forms are identical to those used in the earlier validation examples (Reference 4). The model takes the following form for two-axis multiloop attitude control:

1. Assume a linear model on each axis with a gain, lead (for Class IV aircraft, 0.5 sec), and delay of 0.3 sec (this includes neuromuscular lag which can be modeled separately if desired).
2. Assume continuous control in single-axis tasks, split control in multi-axis tasks. For a two-axis task, the pilot model will switch from one axis to the other based on testing the urgency functions:

P axis controlled if and only if $U_P > U_R$

3. The urgency functions take the form

$$U_P = \left| \alpha \left| \phi_e \right| + \beta \frac{\phi_e}{\left| \phi_e \right|} \dot{\phi}_e \right| \quad (12)$$

$$U_R = \left| \mu \left| \theta_e \right| + \nu \frac{\theta_e}{\left| \theta_e \right|} \dot{\theta}_e \right|$$

where β is equal to zero in single loop multi-axis tasks (this is the case in the two-axis turbulence problem).

4. The quantity α and its counterpart μ on the other axis are determined by the display. (This is the hypothesis mentioned above: that subjective and objective urgency must agree.)
5. Given the above, the model is optimized in the following manner:
 - a. Optimize each pilot model gain for a continuous single-axis task.
 - b. Use these gains to optimize the radial error $r(\phi, \theta)$

$$r(\phi, \theta) = \sqrt{(\alpha\phi_e)^2 + (\mu\theta_e)^2} \quad (13)$$

by perturbing β and its counterpart ν .

- c. The gains should again be checked since it often occurs that optimum multi-axis gains are somewhat lower than single-axis optimum gains.
6. With this procedure complete, the model can be exercised to produce time histories of all dynamic quantities, all statistics of mean, r.m.s, control episode periods, dwell fractions, and other parameters which serve as predictions of the performance of an asymptotically trained human controller for any system disturbance to be modeled such as pilot perception threshold, turbulence, or commands, both continuous and discrete, Gaussian and non-Gaussian.

For the two-axis attitude stabilization task in turbulence, the above algorithm simply requires optimizing two quantities (the pilot model gains) in the separate continuous single axis tasks. Hence

$$\begin{aligned} \text{Longitudinal Urgency: } U_P &= \mu |\theta| \\ \text{Lateral Urgency: } U_R &= \alpha |\phi| \end{aligned} \quad (14)$$

Before examining the quantitative agreement of the model with the flight simulation, certain qualitative similarities with the data will be discussed. Although there has been a wide and rapidly growing recognition that the human pilot behaves as a time sharing device, the sharp division between lateral and longitudinal control has not often been directly exhibited. Two types of examples will be shown. The first illustration is a trace taken during the simulation of lateral versus longitudinal stick deflection, Figure 5. Notice that nearly all stick motion is parallel to either the lateral or the longitudinal control axes. Also note that these traces show that the pilot spent most of his time performing lateral corrections (which is consistent with the dwell fractions computed for this μ/α ratio of 8/1, approximately 0.75 lateral control). The second and most telling illustration of the pilot's switching from one control axis to the other is shown in Figure 6. Here the actual strip chart record is reproduced for a twenty-second sample of the two-axis task. The control episodes have been marked with a step function showing the pilot's apparent control shifting. If the reader will follow the entire record starting at time zero, the left hand side, he can compare the alternation of the control as indicated. Furthermore, if the ϕ and θ time series are compared, it is clear why control shifting takes place when it does.

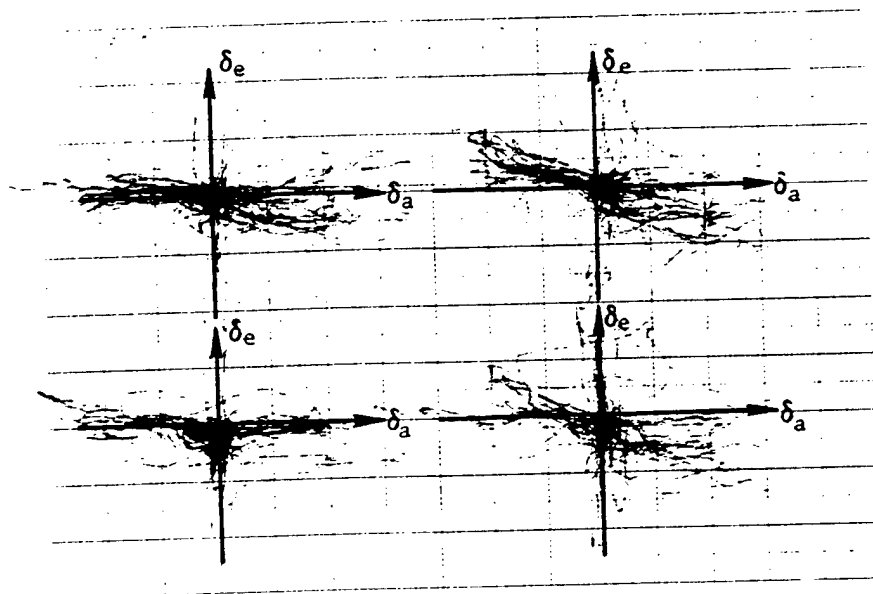


FIGURE 5. LONGITUDINAL VERSUS LATERAL STICK DEFLECTION

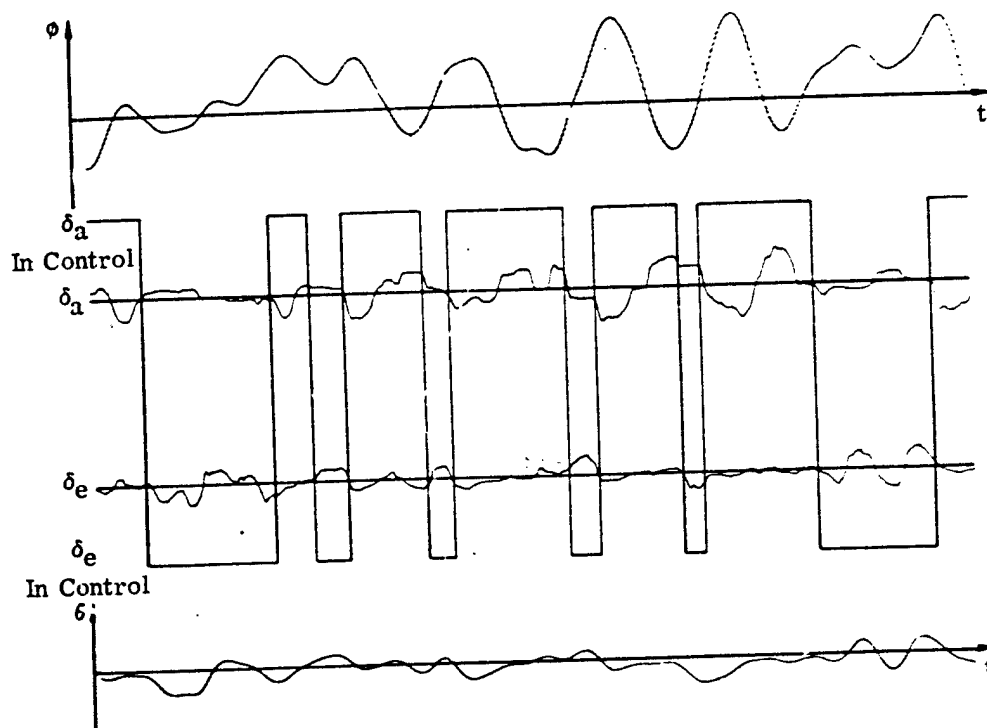


FIGURE 6. CONTROL SHIFTING EPISODES IDENTIFIED FROM SIMULATOR DATA

The most interesting comparisons are, of course, with the quantitative measures of the tracking performance. In the following data presentation, all tracking errors have been normalized to gust levels of 10 ft/sec. This turbulence level is approximately that used in both the analysis and the simulation, but statistical fluctuations produce variations in the actual turbulence level for any given data run, in both the model and the flight simulation.

The first data to be considered are an overall comparison of tracking error predictions and the simulation data. Figure 7 shows a plot of longitudinal θ simulation data versus model predictions. The simulator data are averaged from the 30-second tests for continuous, $\mu/\alpha = 8$, and $\mu/\alpha = 16$ tasks, while the model data were obtained from time histories spanning 1000 seconds. Figure 8 gives a similar comparison for the lateral ϕ data. The agreement here is slightly less good for the $\mu/\alpha = 16$ cases owing to experimental display constraints; at this urgency ratio, the dot motions sometimes went off the display. In particular, note the extremely accurate predictions for the single-axis task (squares). Since the pilot was instructed to minimize the rms distance of the dot from the center of the display, it is of value to compare the radial error data. This is shown for the two-axis data in Figure 9. It should be remarked upon that this agreement between model predictions and simulation data for a two-axis task rivals currently reported results in many single axis studies.

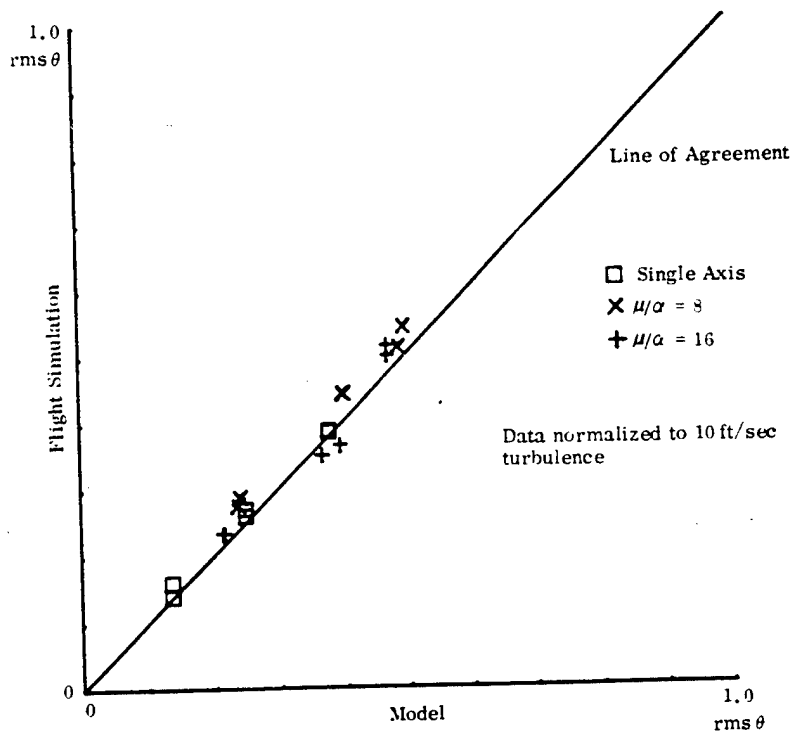


FIGURE 7. MODEL VERSUS FLIGHT SIMULATION DATA FOR ALL LONGITUDINAL TASKS

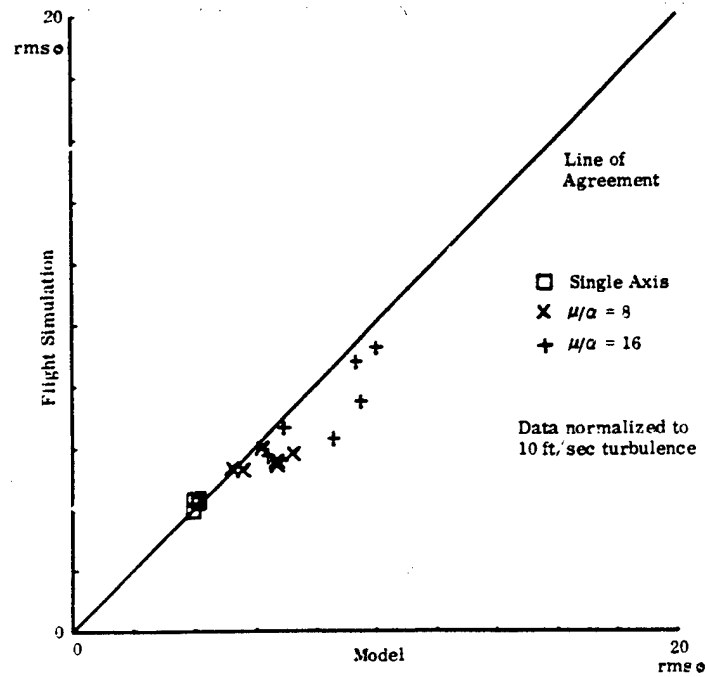


FIGURE 8. MODEL VERSUS FLIGHT SIMULATION DATA FOR ALL LATERAL TASKS

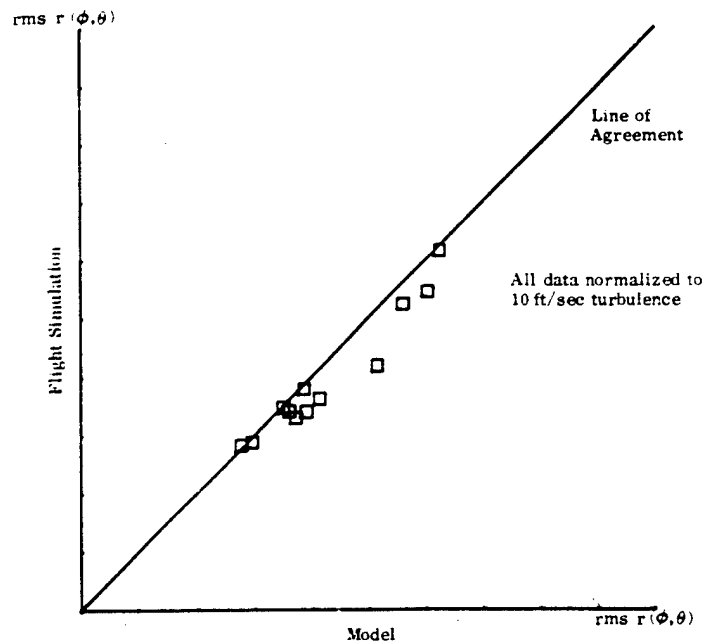


FIGURE 9. RADIAL ERROR OF MODEL VERSUS FLIGHT SIMULATION FOR TWO-AXIS ATTITUDE STABILIZATION

It is useful to examine the individual configurations. A particular type of diagram was evolved that portrays the two axis and the urgency ratio effects. An explanatory example is given in Figure 10. The vertical axis is lateral tracking error, and the horizontal axis is longitudinal tracking error. In order to see how a particular μ/α ratio compares with the open-loop aircraft response and with the continuous single-axis tracking performance, these values have been drawn as lines of constant-tracking error on the diagram as, for example, the line labeled "open loop ϕ ." Thus the two-axis performance would be expected to fall somewhere in the rectangle bounded by these lines. For an urgency ratio of zero, i.e., continuous lateral tracking, the model prediction is the lower right-hand corner of the rectangle. The predictions for ratios of 8 and 16 are shown and labeled with the μ/α ratio. The averaged model data are also shown and labeled.

By referring back to the matrix of lateral and longitudinal dynamics, Figure 4, it is possible to see to what extent lateral and longitudinal dynamics interfere with each other through the pilot. The most striking example, and this is one of the most significant findings of the study, is furnished by a comparison of 1B and 3B, Figures 11 and 12. Both of these configurations have the same lateral dynamics, B, but differ longitudinally. A comparison of the data shows that in 3B, the two-axis control stays within the rectangle for urgency ratios of 8 and 16. However, in 1B the different longitudinal dynamics result in deteriorated tracking at a ratio of 8, and grossly deteriorated tracking at 16. The result of this is that task interference and dynamic mismatching are important aspects of two-axis handling qualities that are seen to exist experimentally, and are subjects that can be accurately predicted and evaluated by the two-axis urgency function pilot model.

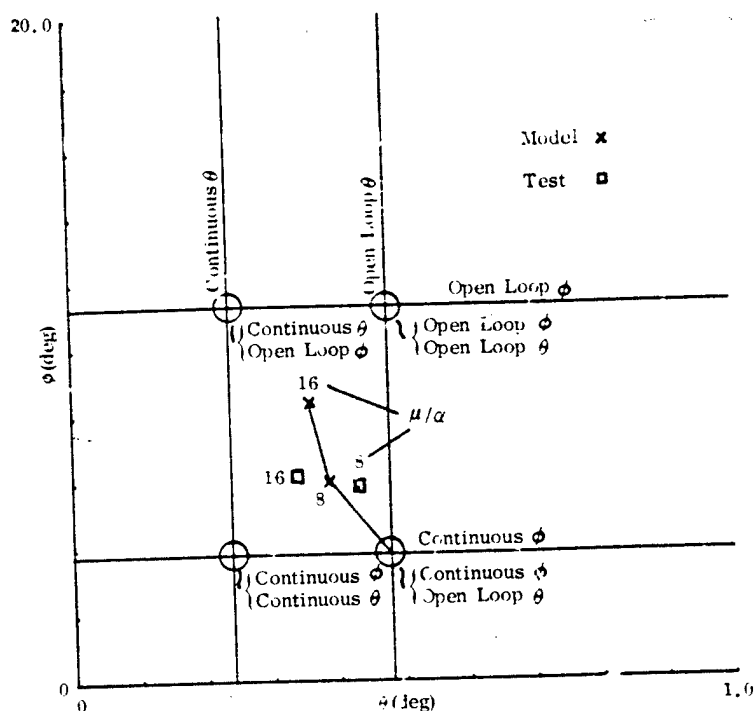


FIGURE 10. HOW TO READ TWO-AXIS DATA PLATES (EXAMPLE)

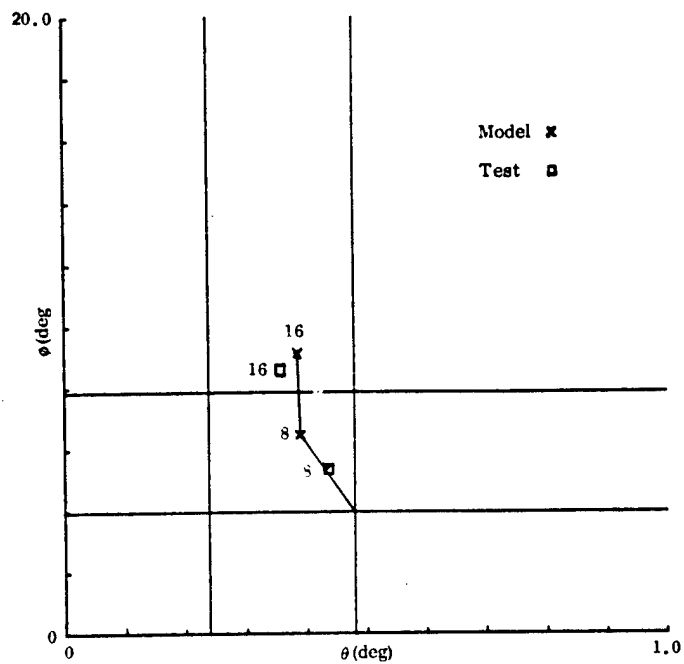


FIGURE 11. MODEL AND TEST DATA FOR F-5 1B

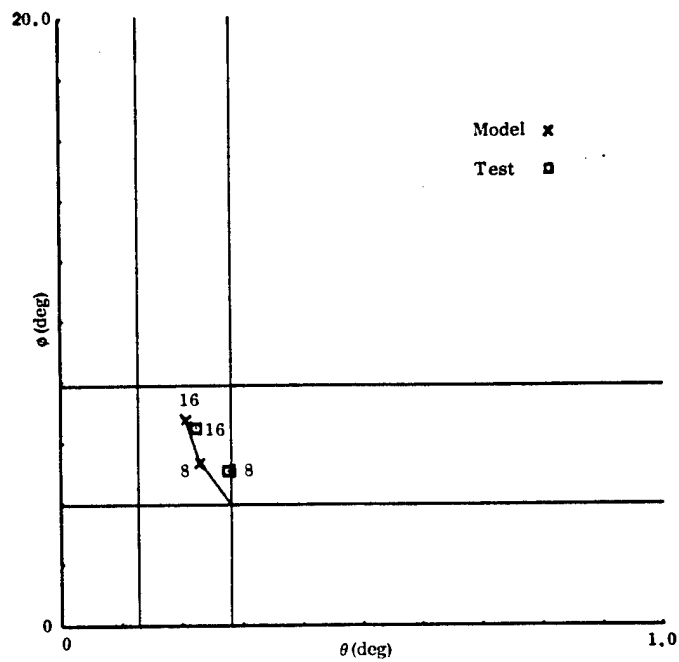


FIGURE 12. MODEL AND TEST DATA FOR F-5 3B

A further point to observe is that even in the configurations where the lateral tracking error is diminished in the $\mu/\alpha = 16$ tasks, the data still fall along the lines of model predictions regarded as interpolations of the urgency ratios. This indicates that when the display limit forced additional attention to the lateral task, the pilot simply shifted his subjective urgency, i. e., his actual weighting of ϕ and θ , thus displacing the data along the urgency interpolation curves. In the other cases where the display limit was not encountered, the data fall close to the curves. This indicates two things: First, the pilot's subjective urgency and the objective urgency of the actual display gains closely agree; second, the greatest source of model prediction inaccuracy is a failure of these two urgencies precisely to match up owing to experimental constraints not modeled in this example.

There is one further statistical comparison that has quite important consequences to aircraft design and air safety, namely, the variation in tracking error as shown in the performance data dispersion of the 30-second simulations. The following Figures 13 and 14 show these 30-second data for two-axis flight simulation data as well as 30-second model data for a comparable number of flights. The significant feature of two-axis handling qualities that these figures exhibit is the propensity of certain configurations to large excursions and tracking error, while others appear more uniform. This demonstrates a varying sensitivity to fluctuations in the effective turbulence power spectrum and probability density during the finite duration flight histories. Here, the model and the flight simulation show a large dispersion, with many flights performing much worse laterally than open-loop response. These

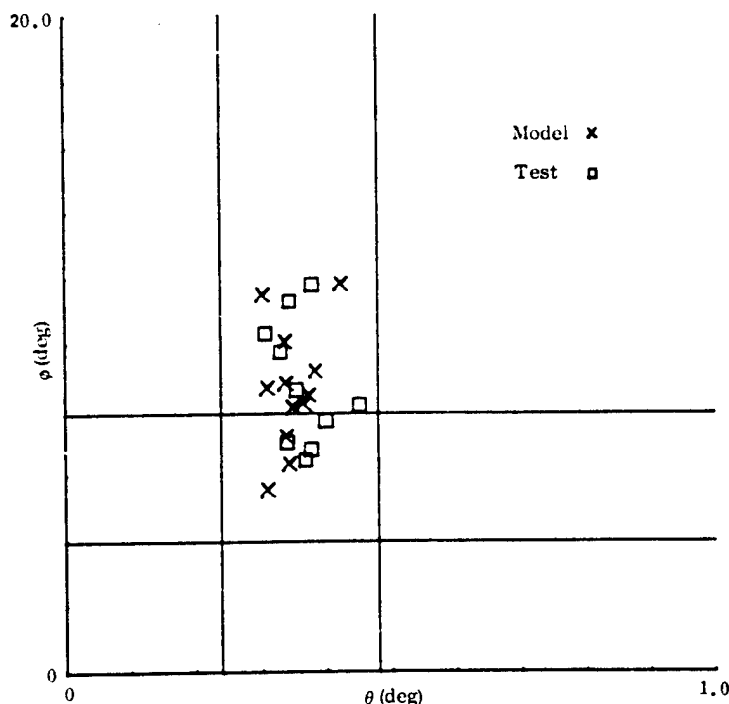


FIGURE 13. MODEL AND SIMULATION DATA DISPERSION FOR 30 SEC FLIGHTS
F-5 1B $\mu/\alpha = 16$

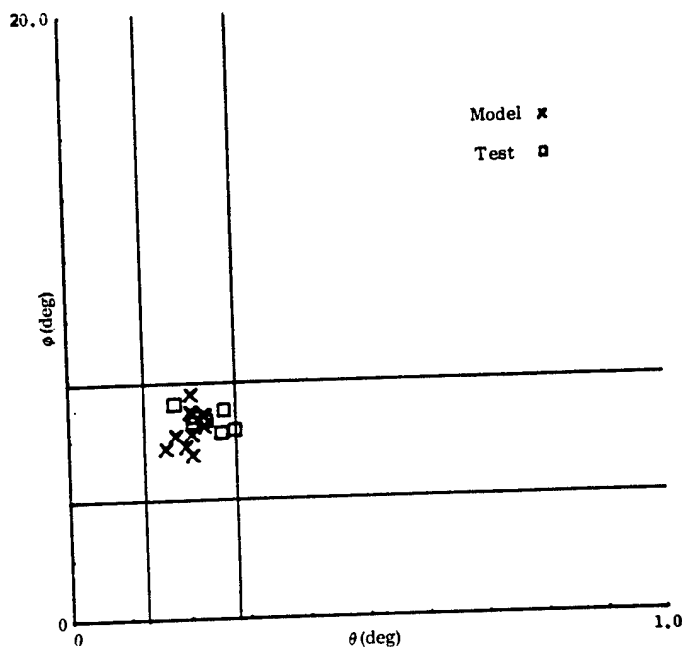


FIGURE 14. MODEL AND SIMULATION DATA DISPERSION FOR 30 SEC FLIGHTS
F-5 3B $\mu/\alpha = 16$

effects indicate that not only do certain combinations of lateral and longitudinal dynamics lead to deteriorated tracking errors, but to a large scatter in the performance as well. The actual attitude excursion encountered in examples such as 1B could, in an actual flight, be severe enough to lead to loss of control. This effect is totally due to the two-axis time sharing of the human controller, since such sensitivity is not observed in the continuous single-axis data, or in other configurations involving the lateral or longitudinal dynamics that together produce these examples. Considering 1B further, Figure 15 shows the model time history of 225 seconds lateral tracking error for continuous single-axis tracking and for two-axis tracking with an urgency ratio of 16. In the continuous case, there is a total absence of large excursions, while in the two-axis history large fluctuations occur in local tracking performance as well as numerous severe sudden excursions.

TWO-AXIS ATTITUDE STABILIZATION OF THE YF-17 IN TURBULENCE

In April 1975, the Northrop multi-axis pilot model was further validated through flight simulation of the YF-17 using the fully general aircraft model on the Northrop LAS/WAVS flight simulator. There were three objectives of this study:

- Demonstrate the generality of the Northrop pilot-aircraft model
- Validate the model for external visual flying
- Evaluate the importance of modeling visual deadbands.

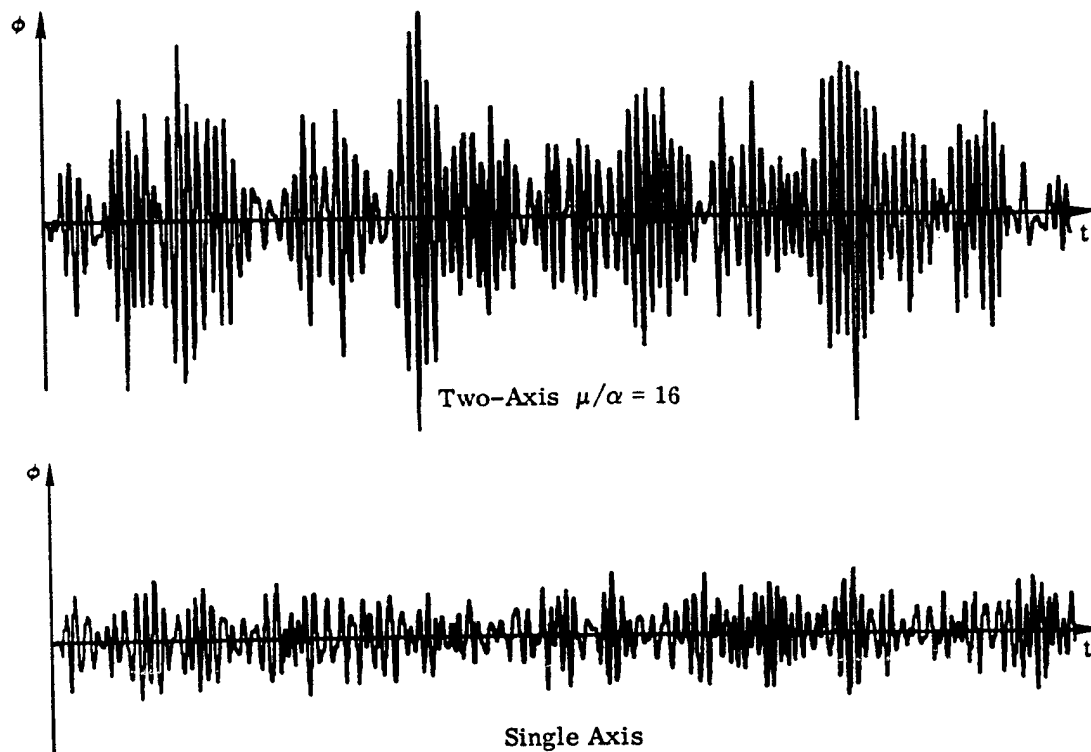


FIGURE 15. LATERAL TIME HISTORIES FOR SINGLE AND TWO-AXIS TURBULENCE TRACKING OF B LATERAL DYNAMICS

The aircraft model included all nonlinear aerodynamics terms and parametric table look-ups as well as the nonlinear control system including automatic flap setting. The cockpit was used with no change from the YF-17 configuration, including stick gradients and all functioning instruments. Dryden turbulence was generated for lateral v and longitudinal w gusts, and stabilization flights were of thirty-second duration.

In order to incorporate visual deadband in a realistic manner, the wide-angle visual-system earth-sky projector projected a horizon of mountainous terrain. By having the pilot (a former U.S. Navy test pilot) attempt to hold zero roll angle, the presence of visual deadband could be both demonstrated and measured from the strip chart recordings. This deadband then was mechanized into the pilot model as an inequality test, not as injected white noise. In this way, the model of the deadband preserved the dynamic characteristics of the actual flight simulation as well as led to the correct statistics.

The model of the piloted YF-17 was exercised by inserting the pilot model into the simulator computers, a small modification. Flights simulations were completed for single- and two-axis tasks at six flight conditions as shown in Figure 16.

	ALTITUDE	5K	10K	30K	40K	60K
MACH						
0.4		X		X		
0.8		X			X	
0.9						X
1.1			X			

FIGURE 16. YF-17 FLIGHT CONDITIONS STUDIED

Plots of model versus flight simulation for the single-axis tasks are given in Figures 17 and 18, where the lateral model data have been completed using the deadband, 3 degrees. The data for the two-axis tasks are shown in Figures 19 and 20. The close agreement of these data between the model and the flight simulation validates the model, including the visual deadband. Computations of the model without the deadband agree poorly for the flight conditions of high dynamic pressure where the

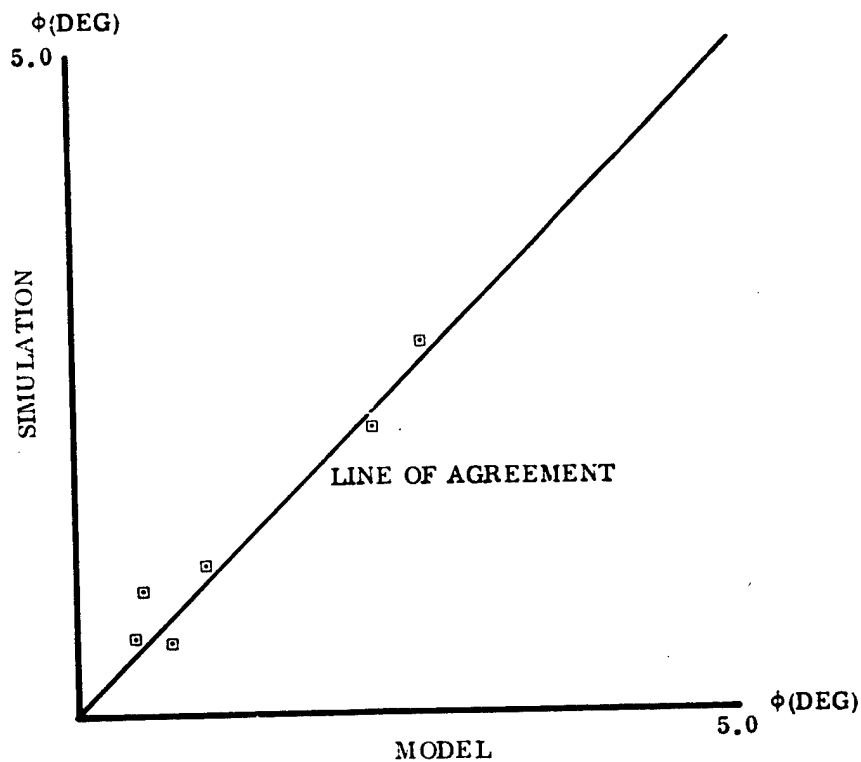


FIGURE 17. LATERAL TURBULENCE TRACKING OF THE YF-17 SINGLE AXIS

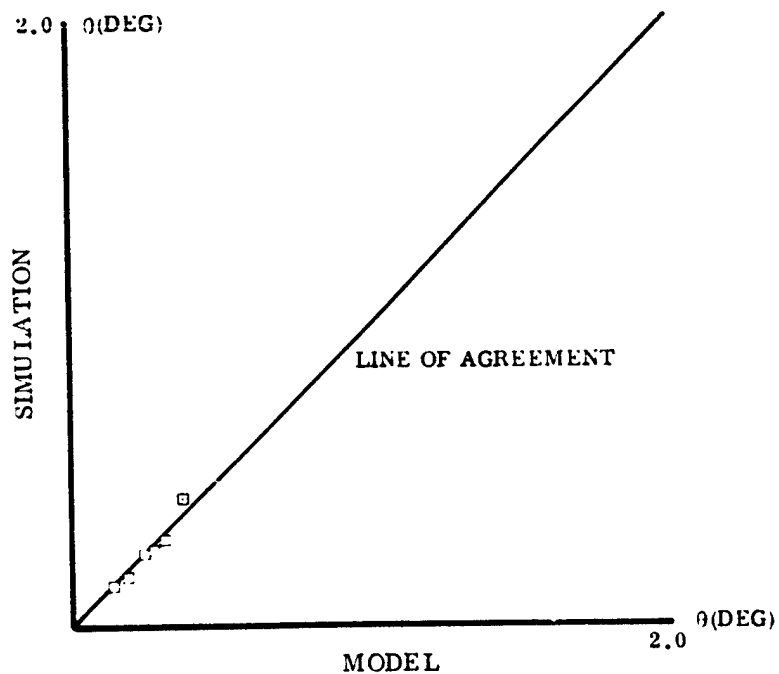


FIGURE 18. LONGITUDINAL TURBULENCE TRACKING OF THE YF-17 SINGLE-AXIS

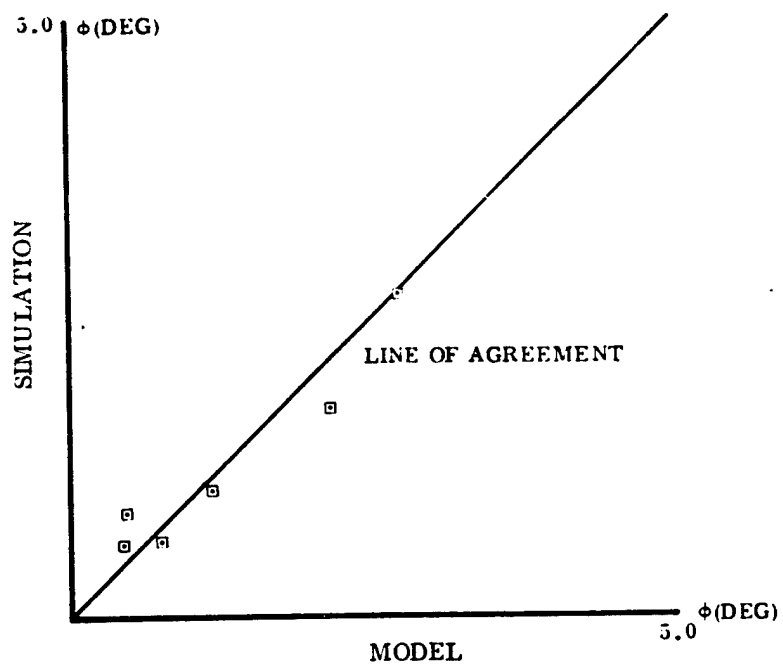


FIGURE 19. LATERAL TURBULENCE TRACKING OF THE YF-17 TWO-AXIS

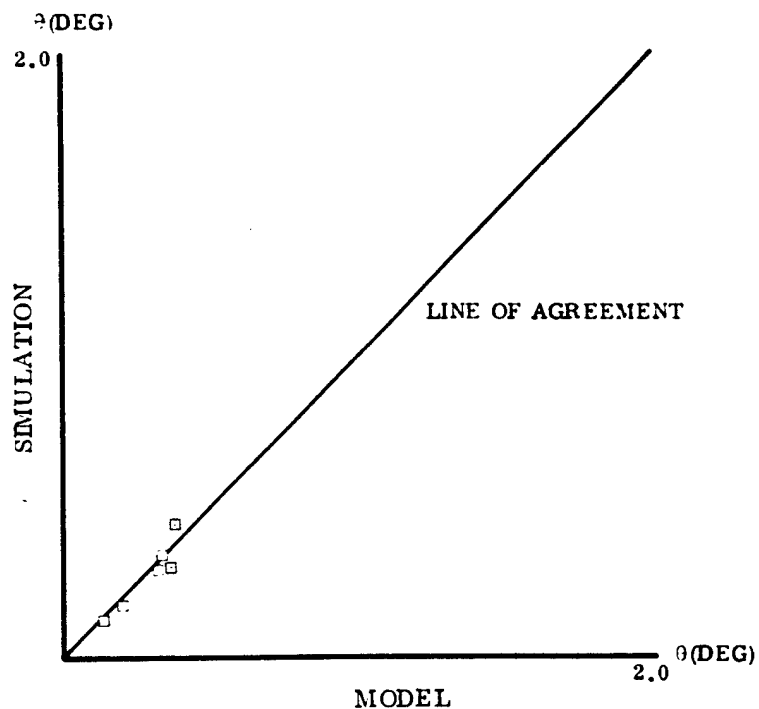


FIGURE 20. LONGITUDINAL TURBULENCE TRACKING OF THE YF-17 TWO-AXIS

turbulence disturbances are low, being roughly one-fifth to one-quarter of the dead-band model values. The effects of dispersion between the data of the thirty-second flight identified in the study of the F-5 in turbulence also show in the flight conditions of the YF-17 at low dynamic pressure. The data from the individual flights are presented for the six flight conditions in Figures 21 through 26.

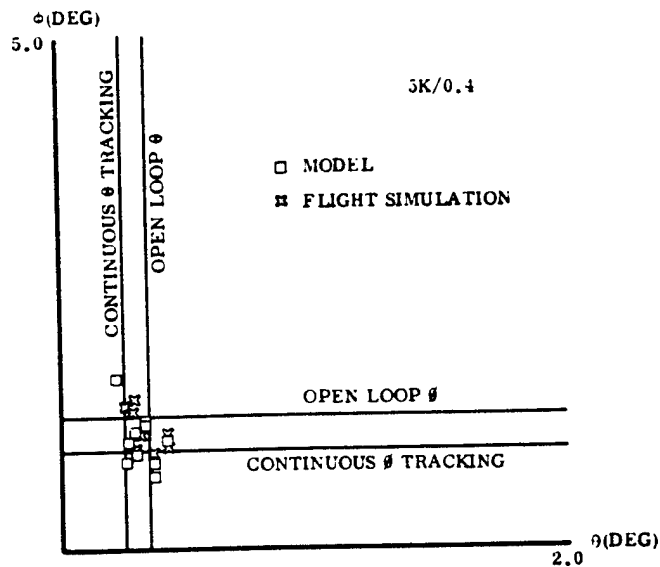


FIGURE 21. TWO-AXIS TRACKING DATA FOR 5,000 FT AT MACH 0.4

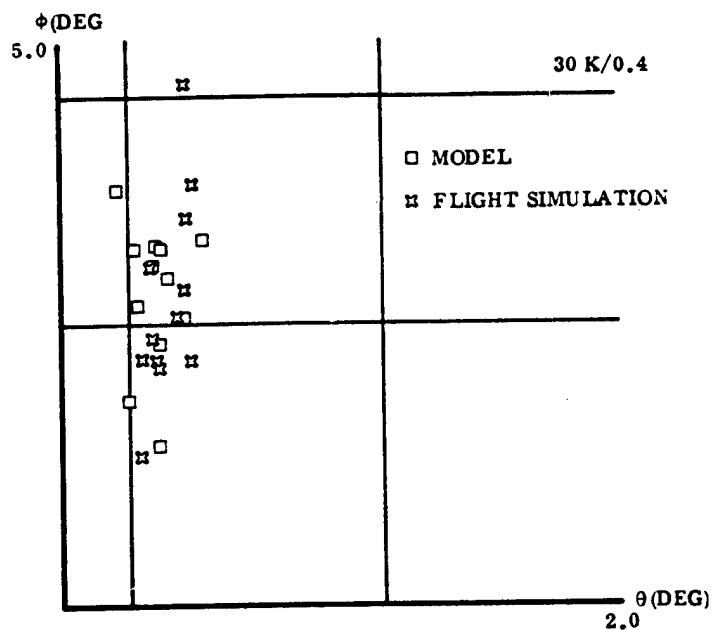


FIGURE 22. TWO-AXIS TRACKING DATA FOR 30,000 FT AT MACH 0.4

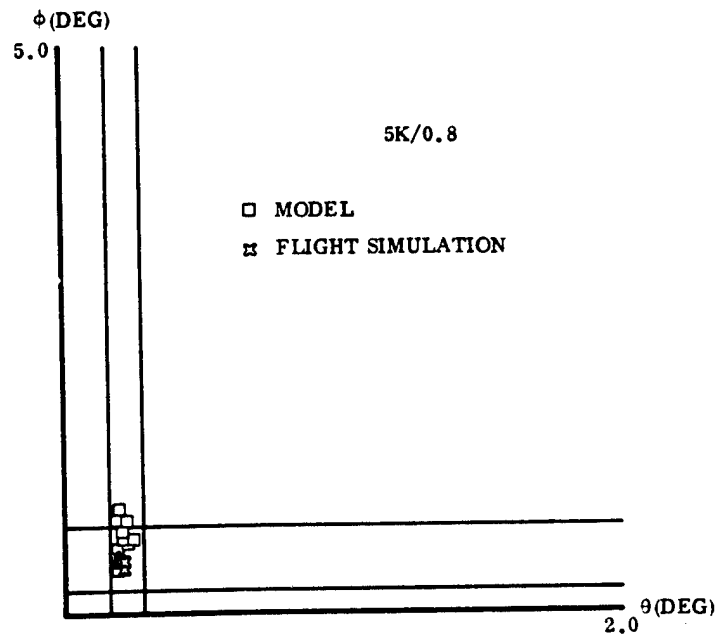


FIGURE 23. TWO-AXIS TRACKING DATA FOR 5,000 FT AT MACH 0.8

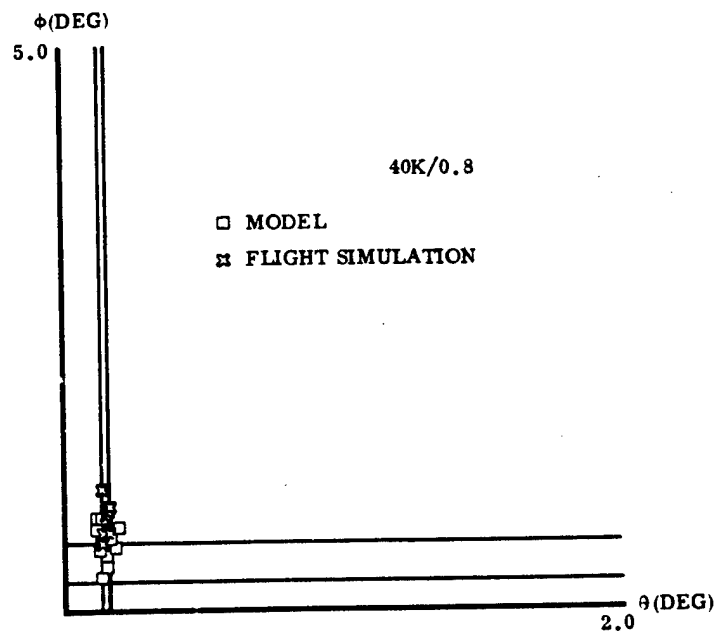


FIGURE 24. TWO-AXIS TRACKING DATA FOR 40,000 FT AT MACH 0.8

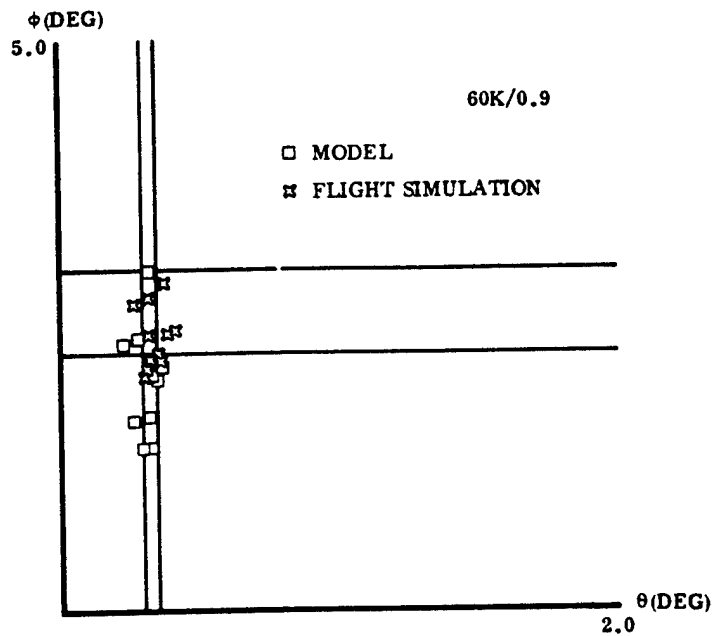


FIGURE 25. TWO-AXIS TRACKING DATA FOR 60,000 FT AT MACH 0.9

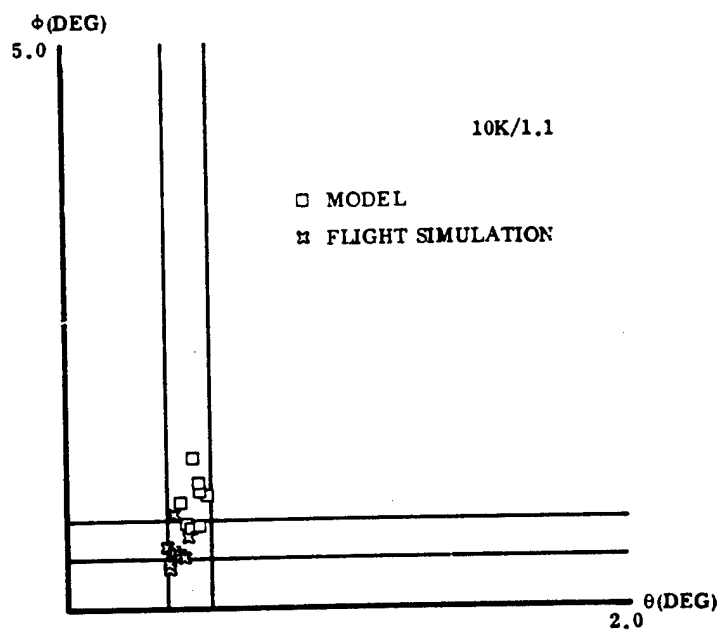


FIGURE 26. TWO-AXIS TRACKING DATA FOR 10,000 FT AT MACH 1.1

APPLICATIONS

Since these methods of digital simulation and multi-axis pilot models allow large scale maneuvers to be studied as the closed loop nonlinear and time-varying problems that they are, there are a large number of applications which can be undertaken. Northrop is currently working on several major examples, and further work will include the following problem areas.

Weapon Delivery

Both air-to-air and air-to-ground problems can be studied. In dive bombing, for example, the analysis includes target acquisition, roll-in, acquisition of the glide slope, tracking using the weapon sight, intense but briefly unstable attitude rate stabilization prior to weapon release, and pull up. The entire maneuver can be digitally "flown" many times in the presence of the realistic environment of low level turbulence and pilot induced system noise, and the statistics of impact error produced.

Landing Approach

The aerodynamics of landing approach are not well represented by constant coefficient linear uncoupled dynamics. Furthermore, the landing task in the presence of low level turbulence make the use of the exact nonlinear models of the problem necessary. Much has been learned from linear methods about the form of the control strategy that the pilot must use, and the study of the behavior of the multi-axis pilot model can utilize this information to produce a comprehensive analysis of the entire maneuver.

Coordination of Analysis and Flight Simulation

The multi-axis pilot model technologies developed at Northrop will not replace piloted flight simulation, but in fact can be used to greatly improve its efficiency. Not only will consistent analytical and experimental results give greater credibility to both, but the ability to screen away needless experiments by analysis will also improve the overall efficiency of aircraft evaluation and design.

Loss of Control at High Angles of Attack

Once aerodynamic descriptions of this flight condition are determined, the nonlinear and time-varying modeling capabilities of the Northrop methods can be immediately applied. The multi-axis pilot models required will be similar to the ones developed for attitude stabilization. As an example, consider Figure 27. This is the time history of one axis of a two-axis attitude stabilization problem. The system remained well-behaved for 120 seconds at which time task interference effects coupled through the pilot model caused a sudden divergence of the system in .3 seconds.

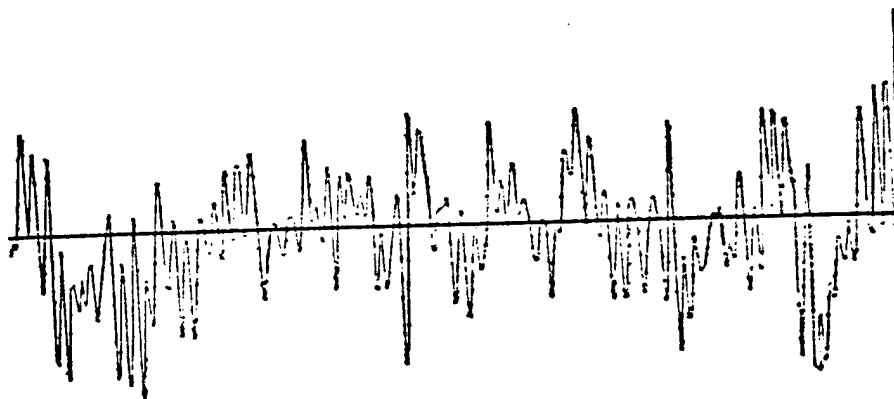


FIGURE 27. SUDDEN LOSS OF CONTROL IN A TWO-AXIS TASK

FUTURE RESEARCH

Apart from continuing work in the areas of the applications discussed above, Northrop is pursuing further development of the basic methods of dynamically modeling human factor pilot model components. The current work addresses the items indicated in Figure 28, which is regarded as the complete pilot model.

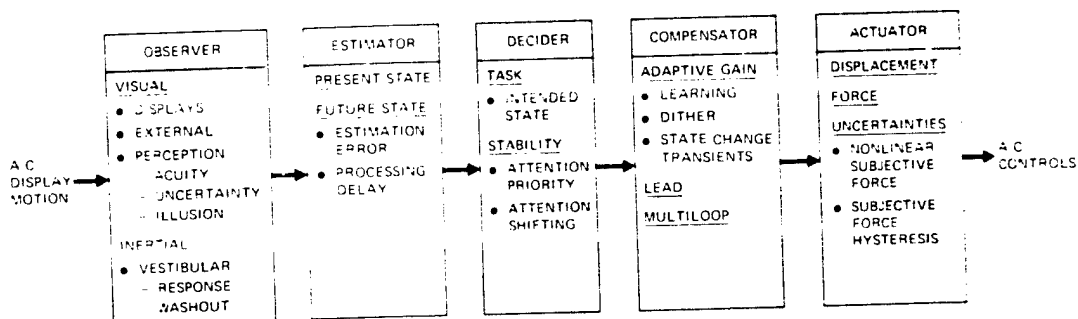


FIGURE 28. TOTAL PILOT MODEL

These five areas of observation, estimation, decision, compensation, and actuation all contribute significantly to the total pilot-aircraft interface. Since only the compensation block lends itself to linear analysis, describing function and optimal control approaches to pilot modeling have been mainly confined to this area. The other functions have been traditionally relegated to a lumped method of degrading the model performance to emulate deteriorated tracking scores through the injection of noise to the model's output. The Northrop time domain pilot model, on the other hand, allows all relevant nonlinearities and time dependent functions of the pilot to be directly represented. The further development of decision algorithms to model attention shifting behavior is currently being performed for several combinations of multi-axis dynamics. Estimation and observation include both the evaluation of

perception position and rate uncertainties and thresholds. Adaptive algorithms to represent the gain optimizing capability of the pilot are also under investigation.

CONCLUDING REMARKS

A comprehensive context for flying qualities prediction and evaluation has been evolved in order to achieve sufficient generality of both pilot and aircraft. Based on the methods of digital simulation, the central part of this technology is the development of multi-axis time domain pilot models which include not only the features of dynamical compensation and human factors data relating to human observation and estimation, but decision and evaluation capability as well.

In order to validate the decision models for attitude stabilization in turbulence, analytical and experimental research was completed using fixed and moving base flight simulation of F-5 and YF-17 aircraft. The pilot-vehicle system model was adjusted by optimization either before, or independent from the corresponding flight simulation. In this way, the predictive capability of the method was demonstrated. The time domain statistics predicted by the model closely matched the experimental data including the episodic control shifting characteristics of the multi-axis human controller. Moreover, an important new aspect of multi-axis flying qualities, cross-task resonance, was identified and reproduced analytically.

As a result of this validation study, the importance as well as the practicality of the basic methodology has been demonstrated. General applications are now possible, and basic research into the decision models for piloted control tasks will include weapon delivery, loss of control at high angles of attack, approach and landing, and dynamic pilot-aircraft interface matching of the longitudinal with the lateral-directional dynamics of the aircraft.

REFERENCES

1. Onstott, E. D., and Salmon, E. P., Airplane Flying Qualities in Turbulence, AFFDL-TR-70-143 Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, Feb. 1971.
2. Onstott, E. D. et al, Prediction and Evaluation of Flying Qualities in Turbulence, AFFDL-TR-71-162 Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, Feb. 1972.
3. Onstott, E. D., Prediction and Evaluation of Flying Qualities in Turbulence, AFFDL-TR-72-92, Proceedings of the Eighth Annual Conference on Manual Control.
4. Onstott, E. D., Multi-Axis Pilot-Vehicle Dynamics, Proceedings of the Tenth Annual Conference on Manual Control, AFIT/AFFDL, April 1974.
5. Onstott, E. D., Validation of the Northrop Multi-Axis Pilot Model, Northrop Report NOR 75-61, March 1975.

TELLING A COMPUTER HOW A HUMAN HAS ALLOCATED HIS
ATTENTION BETWEEN CONTROL AND MONITORING TASKS*

Kenneth D. Enstrom

John Deere and Company
Moline, Illinois 61265

William B. Rouse

University of Illinois
Urbana, Illinois 61801

SUMMARY

The computer's knowledge of how the human has allocated his attention is posed as an important issue in the design of human-computer systems where the two decision makers have overlapping responsibilities. It is argued that it is inappropriate to require the human to continually tell the computer how he has allocated his attention. Instead, a computer algorithm employing fading-memory system identification and linear discriminant analysis is proposed for real time detection of human shifts of attention in a control and monitoring situation. Experimental results are presented that validate the usefulness of the method.

INTRODUCTION

While many processes can be completely automated and thus, no longer require a human as part of the system, there is a large class of processes where only partial automation is possible and require that the human remain part of the system. Thus, a human-computer system is formed. One of the most important issues in the design of human-computer decision making systems is the allocation of decision making responsibility [1]. In many cases, where the human and computer have overlapping abilities, it seems desirable for them to have overlapping responsibilities [2]. Thus, allocation of decision making responsibility becomes situation dependent in that a particular task is assigned to the decision maker who, at the moment, has the time to allocate to the task in question. The main shortcoming of this approach is the possibility of conflicts in the sense that the two decision makers lose track of what each other is doing which results in their competing to perform tasks. Such competition can have disastrous effects on performance [3].

An intuitive solution to this problem is to have each decision maker tell each other of their actions. This may be reasonable from the computer's point of view, but such a dialogue might significantly increase the human's workload and subvert the main reason for utilizing the computer. Thus, we need some method for determining what the human is doing without continually asking him.

*Supported by the U. S. Air Force Systems Command under Contract F33615-73-C-1238.

This paper considers this problem in the context of a control and monitoring situation. As an example, in an aircraft, the pilot's control task is maintaining attitude, altitude, heading and speed while his monitoring task involves maintaining his knowledge of the state of his numerous subsystems (i.e., electric, hydraulic, etc.). Assuming a computer system was available for aiding the pilot in detecting and perhaps correcting changes in subsystem states, then the computer would require a knowledge of the pilot's perception of sub-system states. Otherwise, they might individually initiate actions that would jointly be counterproductive or perhaps disastrous.

We could easily develop other examples for various human-vehicle systems. However, we will jump immediately to the essence of the problem. The basic question is: How can the computer determine that the human has allocated to his monitoring task a portion of his attention in excess of what is needed merely to scan the monitoring situation?

This paper proposes a method for real time detection of how the human has allocated his attention. The method utilizes a fading-memory system identifier in conjunction with a linear discriminant function to detect changes in the dynamics of the control situation that indicate a shift in the human's allocation of attention in excess of what is expected with normal scanning. We will now proceed to discuss the method in detail and consider experimental results that validate the usefulness of the method. The interested reader can find more detail in Enstrom's thesis [4].

APPROACH

Considerable research has been devoted to the study of manual control systems and the identification of the human's dynamics (i.e., transfer function or state model) in control situations [5,6]. We require a method that can operate in real time on a digital computer and successfully identify time-variations in system parameters. A fading-memory least-squares identification algorithm was chosen.

The human plus controlled process were modeled as a discrete linear system using

$$X(k+1) = \Phi X(k) + \Gamma U(k), \quad (1)$$

$$Z(k+1) = H X(k+1). \quad (2)$$

where X is an n state vector, U is a p input vector, Φ is an $n \times n$ state transition matrix, Γ is an $n \times p$ disturbance transition matrix, and H is an $m \times n$ matrix. Following Lee [7], these two equations may be transformed into a canonical form, and then, assuming a single-input, single-output system, into a difference equation of the following form.

$$Z(k) = \sum_{i=1}^n -a_i Z(k-i) + b_i U(k-i) \quad (3)$$

In matrix notation, this equation becomes

$$Z(k) = S(k-1)' \phi \quad (4)$$

where $S(k-1)' = [Z(k-n) \dots Z(k-1) U(k-n) \dots U(k-1)]$ and

$\phi = [-a_n \dots -a_1 \ b_n \dots b_1]$. By minimizing the squared model estimation error, the model parameter vector $\hat{\phi}$ is iteratively estimated using

$$\hat{\phi}(k+1) = \hat{\phi}(k) + D(k)[Z(k+1) - S(k)' \hat{\phi}(k)] \quad (5)$$

where

$$D(k) = P(k) S(k)[S(k)' P(k) S(k) + 1]^{-1} \quad (6)$$

$$P(k) = P(k-1) - D(k-1) S(k-1)' P(k-1) \quad (7)$$

Initially $\hat{\phi}(0) = 0$ while $P(0)$ is usually set equal to an arbitrarily large diagonal matrix.

In order to assure unbiased estimates, the method requires that $\hat{\phi}$ be updated only after every n inputs, where n is the order of the difference equation model. (This avoids biases due to correlated residuals). A further modification reflecting this requirement produces the following set of equations.

$$\hat{\phi}[(k+1)n] = \hat{\phi}(kn) + D(kn)[Z(kn+1) - S(kn)' \hat{\phi}(kn)], \quad (8)$$

where

$$D(kn) = P(kn) S(kn)[S(kn)' P(kn) S(kn) + 1]^{-1}, \quad (9)$$

$$P(kn) = P[k(n-1)] - D[k(n-1)] S[k(n-1)]' P[k(n-1)]. \quad (10)$$

The application of these equations will identify single-input, single-output time invariant systems. However, the method must be modified if systems with time-varying parameters are to be identified.

A slight modification of derivations by Mendel [8] and Morrison [9] allows the identification error vector at time j , $E(j)$, to be multiplied

by a weighting vector at time j , $W(j)$, so that the new squared error term to be minimized is $[W(j)' E(j)]' [W(j)' E(j)]$. The vector $W(j)$ is formulated from the relationship

$$W(j) = \left[\frac{1}{\rho W(j-1)} \right] \quad (11)$$

where ρ is always less than or equal to 1.0. By making ρ less than 1.0, recent errors have a greater influence than old errors in the formulation of new model parameters. Errors are forgotten as the weighting coefficients asymptotically approach zero. Applying these concepts, the final algorithm equations are

$$\hat{\phi}[(k+1)n] = \hat{\phi}(kn) + D(kn)[Z(kn+1) - S(kn)' \hat{\phi}(kn)], \quad (12)$$

where

$$D(kn) = P(kn) S(kn) [S(kn)' P(kn) S(kn) + \rho^{2n}]^{-1} \quad (13)$$

$$P(kn) = \frac{1}{\rho^{2n}} \left[P[k(n-1)] - D[k(n-1)] S[k(n-1)]' P[k(n-1)] \right]. \quad (14)$$

Throughout the remainder of this paper, the term ρ^{2n} will be referred to as the memory coefficient.

Given that we can adaptively identify the dynamics of the human plus controlled process, we need a method of determining when the dynamics reflect shifts of attention. The method chosen was linear discriminant analysis [10].

Assume that we have events that fall into two classes and that for each event j , we have m measurements denoted by y_{ij} , $i = 1, 2, \dots, m$. Then the class to which a particular event belongs can be determined using

$$c_j = \alpha + \beta_1 y_{1j} + \beta_2 y_{2j} + \dots + \beta_m y_{mj} \quad (15)$$

where $c_j = 1$ denotes membership in class 1 while $c_j = 2$ denotes membership in class 2. If we know the classes to which l events belong, then the c_j , $j = 1, 2, \dots, l$ are determined and β_i , $i = 1, 2, \dots, m$ can be determined if $l \geq m$.

First, the measurements are normalized using

$$\tilde{y}_{ij} = y_{ij} - \frac{1}{l_1 + l_2} \sum_{\delta=1}^l y_{i\delta} \quad (16)$$

$$\tilde{c}_j = c_j - \frac{1}{l_1 + l_2} \sum_{\delta=1}^l c_{\delta} \quad (17)$$

where l_i is the number of events which belong to class i . Next a set of simultaneous linear equations are formed

$$\begin{bmatrix} \tilde{y}_{11} & \tilde{y}_{21} & \cdots & \tilde{y}_{m1} \\ \tilde{y}_{12} & \tilde{y}_{22} & & \\ \vdots & & \ddots & \\ \tilde{y}_{1l} & & & \tilde{y}_{ml} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} = \begin{bmatrix} \tilde{c}_1 \\ \tilde{c}_2 \\ \vdots \\ \tilde{c}_l \end{bmatrix} \quad (18)$$

or, more compactly,

$$\tilde{Y} \beta = \tilde{C} . \quad (19)$$

The least-squares solution is the familiar

$$\hat{\beta} = (\tilde{Y}' \tilde{Y})^{-1} \tilde{Y}' \tilde{C} . \quad (20)$$

Tatsuoka discusses issues such as whether or not two classes are significantly different and the determination of the probability of class membership. The interested reader is referred to reference 10.

The above identification/discrimination methodology is proposed for determining how the human has allocated his attention. The following discussion of experiments will detail the specifics of the method's implementation.

EXPERIMENTS

Three experiments were designed to test the proposed identification/discrimination method. While the first experiment tested the scheme's general ability to detect extreme changes in subject tracking dynamics, the

conditions of the second experiment were chosen to represent more realistic situations and test the method over a range of dynamic systems. The last experiment was closely controlled to provide data from a group of well-trained subjects.

An identical testing setup was used in all of the three experiments. A one-dimensional pursuit tracking task was displayed in the vertical direction on a CRT. (See Figure 1). The CRT was a Hewlett Packard model 1310A, with a screen size of 38.1 X 27.9 cm. A PDP-11 was coupled with a custom built graphics display unit to drive the CRT display. A two axis displacement joystick from Measurement Systems Inc. (model 521) provided the means for controller input. Later, when side tasks were added to the tracking task, the subjects' answers were entered via a keyboard unit from an Infoton Vistar II terminal. To simplify the operation of the keyboard, the unneeded keys were covered and the carriage return button, which was used to signal the completion of data entry, was extended for easy location. (See Figure 2.) All of the tracking tests were five minutes in length.

The displayed input symbol was a square of size 0.68 cm per side, and the output symbol was a cross of similar size. (See Figure 3.) The subject was instructed to move the joystick so as to match the cross's position as closely as possible to that of the moving square. The input path was generated by passing gaussian white noise through a second order filter. A sampled-data second order filter ($\Delta t = 0.10$ sec), with a gain of 1.0, natural frequency of 1.0 radian per second, and a damping ratio of 0.5 produced an input path (filter output) with a standard deviation of 447 screen units. The screen coordinates in the vertical direction were -2047 to + 2047. The same input path was used for all tests and the square's position was always updated at a rate of ten times per second.

The joystick range was plus and minus 30 degrees from the vertical position with the full range of the joystick matching the full range of the screen coordinates. The joystick gain was 68.2 screen units per degree; that is a one degree movement in the joystick resulted in an input change of 68.2 screen units. Two stable and two unstable control dynamics were selected for use in the experiments. (See Table 1.) The joystick was of negligible mass and was originally operated without spring resistance. The update frequency for the cross was 10 times per second for experiments one and two but was changed to 20 times per second for experiment three. Also, a light resistance spring was added to the joystick and controller gain was increased from 1.0 to 2.0 for experiment three.

In the first experiment, one subject was tested with two different controller dynamics. Both tests consisted of alternate periods of normal tracking and periods of no tracking, where the joystick location was kept stationary. In the first test, the subject used a type 2 controller (see Table 1), to track for 60 seconds intervals, with dead times between these periods of 10, 20, and 30 seconds. The second test was performed with a type 1 controller. After a 60 second period of continuous tracking, the remainder of the five minute trial consisted of alternate periods of 30 seconds tracking and 5 seconds of dead time.

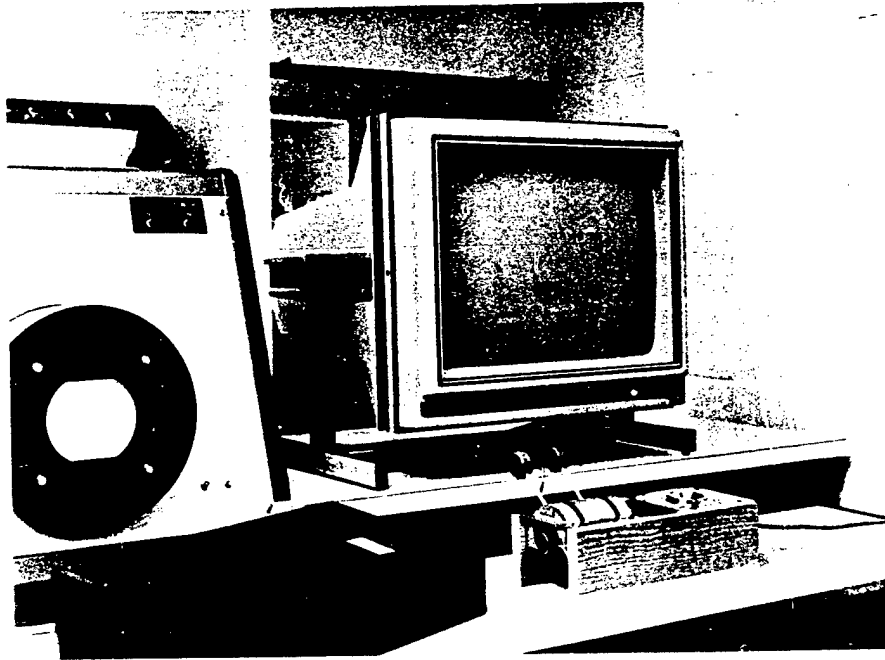


Figure 1. The experimental setup.



Figure 2. The keyboard used for arithmetic sidetasks.

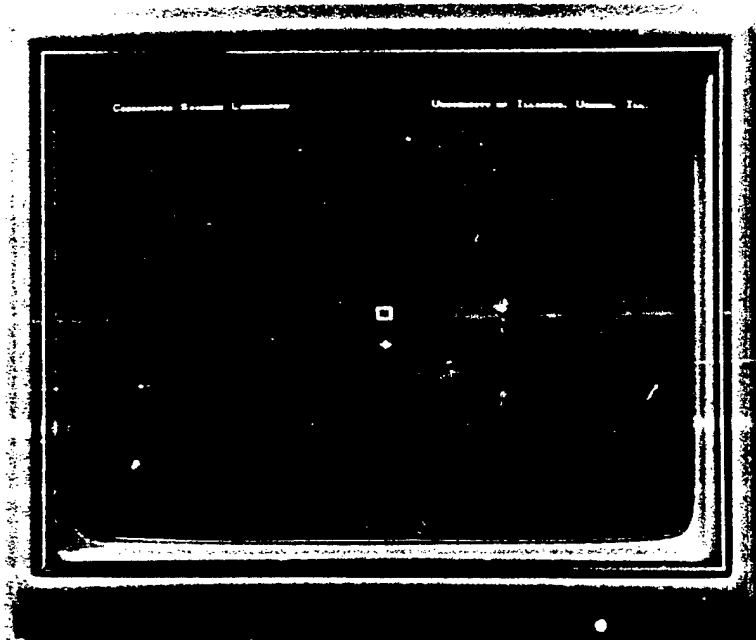


Figure 3. The tracking situation.

TABLE 1
INPUT CONTROL DYNAMICS

	System	Parameters	Sampling Period	Joystick
Type 1	$\frac{K}{(s/\omega_n)^2 + [(2\zeta s/\omega_n) + 1]}$	$K=1 \quad \omega_n=7.8 \quad \zeta=.371$	$\Delta t=0.1$	No Spring
Type 2	$\frac{K}{(s/\omega_n)^2 + [(2\zeta s/\omega_n) + 1]}$	$K=1 \quad \omega_n=1.60 \quad \zeta=.371$	$\Delta t=0.1$	No Spring
Type 3	$\frac{K}{s - 0.5}$	$K = 1$	$\Delta t=0.1$	No Spring
Type 4	$\frac{K}{s - 0.5}$	$K = 2$	$\Delta t=0.05$	Spring

After testing the identification scheme with these extremely large changes in subject control dynamics, a more realistic side task replaced the dead times. An arithmetic problem requiring the multiplication of two numbers was chosen as the secondary task. The subject's answer to the problem was entered on the specially designed keyboard (see Figure 2). To minimize the physical activity associated with the side tasks, the data entries were made without removing eye contact from the CRT screen. Each subject was trained on the keyboard to insure familiarity with its operation before any side task tests. All of the subjects tested were right handed and each controlled the joystick with his right hand and operated the keyboard with his left.

The side task was introduced to the subject by displaying it on the CRT tracking screen. (See Figure 4.) When the side task appeared, the subject had been instructed to compute the answer and enter it on the keyboard at a self-paced rate that would eliminate mistakes. The multiplication problem was displayed until the return key was hit, signalling completion of the data entry. To prevent unreasonably long reaction times or neglect of the main tracking task, subjects were scored on both RMS tracking error and the average reaction time to complete the real side tasks. Peer pressure due to the posting of results appeared to serve as ample subject motivation. To control the difficulty of the side tasks, each multiplication included a two digit number and a one digit number; forming a three digit product less than 250.

The scanning required to detect and interpret a side task would require a short attention shift before mathematical calculations were begun. To determine if the scanning effect resulted in significant changes in the control dynamics, "artificial" side tasks were also used. An artificial side task differed from a real side task by using a "K" rather than an "X" for the multiplication sign. (See Figure 5.) The artificial side tasks would still require visual scanning, but would not necessitate mental attention after the text was recognized as a false command. These tasks were automatically erased after being displayed for five seconds. Three actual side tasks were included in each five minute test - always at times 2, 3, and 4 minutes. Since the subjects of the third experiment were only exposed to two side task runs, it is unlikely that the timing of the tasks would have been anticipated by the subjects. Twelve artificial side tasks were randomly distributed throughout the five minute tracking period, with one restriction: artificial side tasks would always follow real side tasks by a sufficient amount of time to guarantee completion of the real task before the introduction of an artificial one. See Table 2 for a summary of the side task times, types and screen locations.

In the second experiment, two subjects were tested with three sets of control dynamics. For each dynamic type, both subjects were run with and without side tasks. The normal tracking task was run first to familiarize the subject with the system dynamics and allow him to devise a satisfactory operating strategy. Then, within a fifteen minute period, a second test was run with the same control dynamics and the predetermined side tasks. Side tasks with different arithmetic problems were used on successive tests.

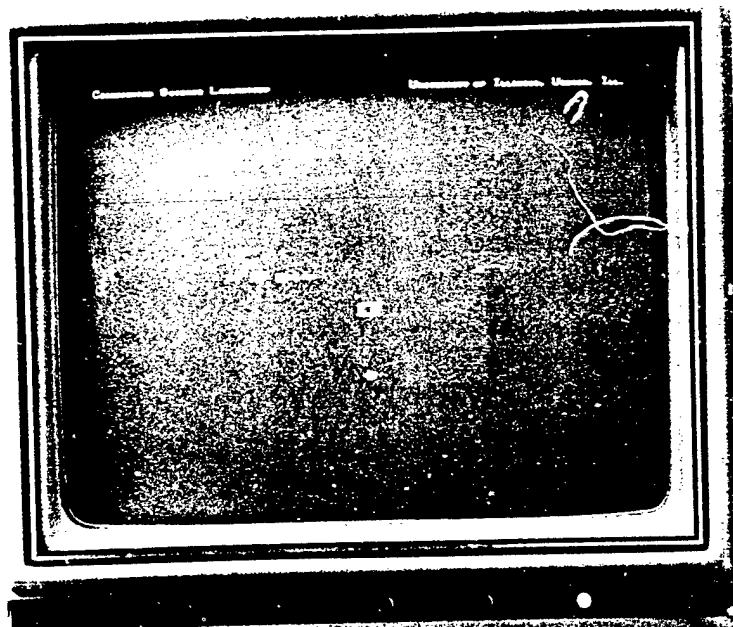


Figure 4. Tracking with a real sidetask.

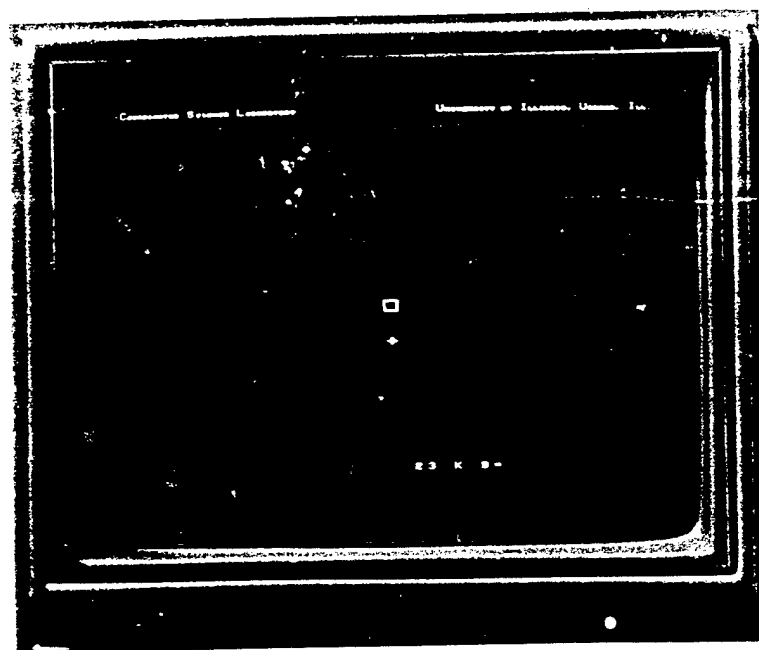


Figure 5. Tracking with an artificial sidetask.

TABLE 2
SIDE TASK INFORMATION

		<u>Number</u>	<u>Type</u>	<u>Position</u>	
				X	Y
5					
	—	4.720	15	K	2200 1500
	—	4.444	14	K	700 1800
4					
	—	4.000	13	X	3000 300
	—	3.860	12	K	3100 3500
	—	3.750	11	K	1200 700
	—	3.470	10	K	2300 2500
3					
	—	3.000	9	X	1000 2200
	—	2.720	8	K	100 1800
	—	2.500	7	K	3000 1500
2					
	—	2.000	6	X	2100 2800
	—	1.720	5	K	1000 1500
	—	1.250	4	K	3000 2000
1					
	—	0.805	3	K	50 50
	—	0.416	2	K	500 3000
	—	0.278	1	K	2200 500
0					

Four subjects were carefully trained on control type 4 for the last experiment. Four practice trials of straight tracking were spaced within a three day period. After the fourth trial, the subjects were tested again with side tasks. The initial reactions to the side tasks were more drastic than anticipated. Several side task errors occurred, while one subject lost control of the output cross, causing the coordinates of the cross to exceed the boundaries of the CRT display area. For this reason, these side task trials were duplicated on the following day with different arithmetic problems. Subject number three was run a third time, due to an error resulting in loss of tracking data.

RESULTS

The raw data collected from each experimental trial consisted of the input and output signals for the tracking tasks and, when side tasks were included in the trial, start and completion times for each task.

Results of the first experiment proved that the identification/discrimination method was capable of adapting to unrealistically large changes in the control system. The two data sets were analyzed using a second order difference equation model and a range of memory coefficients. Substantial changes occurred in the model's four parameters directly corresponding to the dead periods of joystick control. An identification memory coefficient of 0.96 produced model parameter changes that were easily detected by visual inspection of time plots of the parameter values. Since the conditions of this experiment were oversimplified, further analysis was not pursued.

The remaining discussion of side task detection will be divided into two areas. First, the parameters of the identification scheme (i.e., model order and memory coefficient) will be discussed. Then, the application of the linear discriminant function will be considered. The tracking and side task data from experiment two is the basis of these discussions. Finally, we will discuss the results when the data from the third experiment was processed according to the findings from experiment two.

The difference equation model order, which was arbitrarily set to two for experiment one, was studied by forming control models of orders one thru five for the six non-side task tracking data sets from experiment two. The outputs from these models were compared to the real output values to determine an RMS fitting error. The identification model memory coefficient was set to 1.0 for this analysis since side tasks were not used, producing fairly constant system parameters. As expected, the fitting error decreased with increasing model order. Trade-offs between accuracy and feasibility for use in real time applications resulted in a compromised selection of a second order model.

To adaptively identify time-varying system parameters, a memory coefficient less than 1.0 was required. As the coefficient was decreased,

changes in the system would affect this identification scheme quicker. However, noise and inaccuracies would also increase with smaller memory coefficients. Thus, a tradeoff was necessary between identification response time and the accuracy of the results. To determine a suitable memory coefficient, the RMS fitting error was determined for side task data runs from experiment two with memory coefficients of 1.00, 0.95, 0.90, 0.80, and 0.70. The fitting error decreased as the memory coefficient was reduced and then started to increase with a coefficient of 0.70. A value of 0.90 was selected since memory coefficients less than 0.90 produced erratic parameter values - even for non-side-task tracking periods.

Now that a second order model and a memory coefficient of 0.90 have been established, the use of discriminant analysis to detect control system changes will be discussed. A reasonable first choice of discriminant measures was the four parameters of the model, (i.e., the four coefficients of the second-order difference equation). If changes in these parameters were linearly related to side task attention shifts, the discriminant function would be able to detect the side task periods. Visual examination of the parameter vs. time plots indicated that absolute changes in the parameter values did not always correlate well with the side task times. A measure representing the relative changes of the parameters was proposed to correspond more consistently to the side task periods. A long term average of each parameter was calculated and, by comparing the current parameter value to the long term reference value, relative changes were measured. By summing the relative changes over short periods of time, four more discriminant measures were formed. A final measure was added to the group of eight properties by forming an average of the squared tracking errors.

Now we will consider the effects of some of the variables associated with the detection scheme. Such variables include the window size used in the formation of the long term parameter reference values, the relative measure window size, and the a priori probability associated with each classification type. Then, we will consider the scheme's ability to use a single set of sample class measurements to analyze trial data runs within subjects and across dynamics - and within dynamics and across subjects. The scheme's performance with fewer discriminant measures is also of interest.

We have not attempted to determine the optimal settings for the identification/discrimination scheme. Instead, our efforts were aimed at showing some overall trends and characteristics associated with the side task detection algorithm. In many cases when the effect of a single variable was analyzed, the effects of all other variables were assumed to be independent.

Long term reference window sizes of 10.0, 20.0, 25.0, and 30.0 seconds were considered. Fixing the probability of a side task event at 0.05 and a relative measure window size of 6.0 seconds, we found that a reference window size of 25.0 seconds was the only value that detected every side task, and resulted only in false alarms that could be related to artificial

side tasks.

Relative window sizes of 4.0, 6.0, 8.0, 12.0 and 20.0 seconds were considered. Although the results did not show any one window size to be consistently superior to the others, a window size of 8.0 seconds normally produced side task detections with a greater degree of certainty than the others.

The discriminant function requires an initial probability guess for the likelihood of an event being from one of the two classes. As the probability estimate of an event's occurrence is decreased, the discriminant function will predict a lesser number of events for that particular class. Data from experiment two was analyzed using probabilities of a side task event equal to 0.50, 0.10, 0.05, and 0.01. We found that a probability of 0.05 resulted only in false side task detections that were possibly related to an artificial side task, and the one real side task detection that was missed was also missed with values of 0.10 and 0.01.

Until this point, the discriminant function learning samples were formed from the same data file that was to be analyzed. Possibly the similarity among subjects and/or dynamics was great enough to permit the use of one set of discriminant function coefficients in the analysis of different experimental trials. Because experiment two used only two subjects, but three different controller dynamics, this data will be analyzed across the dynamics, but within each subject. Individual learning samples were summed together making one large learning file. The results of these tests are shown in Table 3. Overall, the detection algorithm did not work very well across the different dynamics used in this investigation. However, note that in both subject sets the controller type 2 run was identified far better than the rest. Possibly the discriminant function for this control system was near the average of all three.

Nine measures were used with the discriminant function to take full advantage of as many measures as seemed reasonable. The number of measurements required to make detections was considered by analyzing the same sets of data with a decreasing number of measures. In each reduction, the poorest remaining indicator was dropped. The average order of importance of the measures from experiment two were as follows:

Order of Importance	Measure Number	Description
1	7	Relative Change of Measure No. 3
2	6	Relative Change of Measure No. 2
3	3	Input Value 0.2 Seconds Ago
4	8	Relative Change of Measure No. 4
5	5	Relative Change of Measure No. 1
6	2	Out Value 0.2 Seconds Ago
7	4	Input Value 0.1 Seconds Ago
8	1	Output Value 0.1 Seconds Ago
9	9	Tracking Error Squared

TABLE 3
RESULTS OF USING SAME DISCRIMINANT FUNCTION
ACROSS DYNAMICS, WITHIN SUBJECTS

Subject Number	Controller Type	Number of 2 Second Intervals Detected During Side Task No.			Number of 2 Second Intervals Falsely Detected	
		1	2	3	Total	R.A.S.T.*
1	1	3	0	0	0	0
1	2	6	1	6	0	0
1	3	0	0	1	0	0
2	1	0	2	3	0	0
2	2	6	8	4	0	0
2	3	0	0	0	0	0

* Number related to artificial aide task

TABLE 4
EFFECT OF REDUCING THE NUMBER OF DISCRIMINANT FUNCTION PROPERTIES

Subject Number	1	1	1	2	2	2
Controller Type	1	2	3	1	2	3
Number of Properties	Number of Two Second Intervals Detected During Side Task Number					
	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
9	6 2 0	6 4 8	4 4 5	4 8 3	4 7 5	8 0 5
8	0 0 0	3 1 4	4 5 5	5 8 3	5 6 5	9 0 4
7	0 0 0	3 1 4	4 5 5	4 8 2	5 6 5	9 0 4
6	0 0 0	3 1 4	3 4 3	5 8 2	5 6 5	8 1 4
5	0 0 0	4 0 4	4 4 3	4 8 2	5 6 4	8 1 3
4	0 0 0	5 0 2	0 0 3	4 8 2	5 6 3	3 0 3
3	0 0 0	5 0 1	3 0 3	8 7 0	1 2 1	- - -
2	0 0 0	0 0 0	3 0 3	8 6 0	0 2 1	- - -
1	0 0 0	0 0 0	3 0 3	8 6 0	0 2 0	- - -

Three important points should be noted about the order of the above property list. First, the relative measure of each property was always a better indicator than the absolute property measure. Note also that of the first five important properties, four of them were relative measures. Secondly, the error squared term which was initially thought to be a good indicator was the poorest. This result was not too surprising after reviewing the circumstances. The error is not only a function of the subject's output position, but also a function of the input position of which the subject has no control. The final point is that the most important variable was the relative change of the measure related to the input value of 0.2 seconds ago. Since the normal reaction time delay for a human controller is in this range, we might conjecture that the subject's time delay is affected during shifts of attention. The results of reducing the number of discriminant function measures appear in Table 4. Of the six data sets tested, four of them were still successfully processed using only five properties. Note that the other two data sets contained side tasks that were not detected with the total nine properties.

Now data from the third experiment will be analyzed using the algorithm parameters (i.e., model order, memory coefficient, window sizes, etc.) determined from the data of the second experiment. Summarizing these parameters, the final side task detection algorithm identified four parameters of a second order difference equation model using a fading-memory identification method with a memory coefficient of 0.90. The four model parameters plus a relative measure of each parameter's changes plus a tracking error squared measure were utilized by a linear discriminant function which was "trained" with sample data from each of the two classifications.

The general results of experiment three are shown in Table 5. Note that every real side task was detected, and all other side task predictions were possibly related to artificial side tasks. Thus, the results of experiment two appear to be applicable to at least some additional control systems with different subjects and different controller types.

The data from experiment three was also used to consider the side task detection algorithm's ability to process data across subjects within the same dynamics using a single learning classification sample group. These results appear in Table 6. The method seems to work reasonably well for different subjects using the same dynamics. Notice that only one real side task was not detected, and the only false detections were again related to artificial side tasks. This is somewhat surprising because, the four subjects did not have similar tracking skills - one was much better, while one was consistently poorer.

Finally, the algorithm's ability to operate in real time was examined. Using a DEC-10 computer, 3.917 cpu seconds were required to initialize variables and read in the raw tracking data from a five minute trial. The time required to form the eight measurements (the squared error term was not used), and make predictions with predetermined discriminant coefficients was 6.333 cpu seconds. Real time use appears to be highly supported, since only about ten cpu seconds were used to process a total of five

TABLE 5
GENERAL RESULTS OF EXPERIMENT 3

Subject Number	Controller Type	Number of 2 Second Intervals Detected During Side Task No.			Number of 2 Second Intervals Falsely Detected	
		1	2	3	Total	R.A.S.T*
3	4	6	5	6	0	0
4	4	6	4	10	2	2
5	4	6	4	3	2	2
6	4	5	11	8	0	0

*Number related to artificial side task

TABLE 6
RESULTS OF USING SAME DISCRIMINANT FUNCTION
FOR ALL SUBJECTS IN EXPERIMENT THREE

Subject Number	Controller Type	Number of 2 Second Intervals Detected During Side Task No.			Number of 2 Second Intervals Falsely Detected	
		1	2	3	Total	R.A.S.T*
3	4	6	3	5	3	3
4	4	4	0	9	0	0
5	4	6	4	2	3	3
6	4	5	7	5	0	0

*Number related to artificial side task

minutes of tracking data. Assuming a PDP-11 to be five to ten times slower than a DEC-10, a PDP-11 or other similar sized computer could also be used for real time application of the detection algorithm.

CONCLUSIONS

If a human and computer are to successfully interact in situations where they have overlapping responsibilities, they need to know what each other is doing. However, it is unreasonable to require the human to continually inform the computer of his actions. A method is needed for the computer to detect what the human is doing.

This paper has considered human-computer interaction in control and monitoring situations and proposed a method utilizing a fading-memory system identifier and linear discriminant analysis that allows real time prediction of how the human has allocated his attention between the control and monitoring tasks. Experimental results have validated the usefulness of the method.

To actually implement the method in, for example, an aircraft or spacecraft, further research is needed to develop heuristics for handling problems such as periods when the human completely stops tracking. Multi-class discriminant analysis and perhaps some nonlinear method might be appropriate to this problem.

The proposed method may also be useful as a research tool for those investigating motor skills or attention allocation. Such research would be of use in interpreting the outputs of multi-class discriminant analyses.

To conclude, we have posed what is a very real problem in the design of human-computer systems. Our solution, in the context of control and monitoring situations, appears feasible in a real time setting, but requires more development work before actual implementation.

REFERENCES

1. W. B. Rouse, "Design of Man-Computer Interfaces for On-Line Interactive Systems," Proceedings of the IEEE, Special Issue on Interactive Computer Systems, Vol. 63, No. 6, pp. 847-857, June 1975.
2. W. B. Rouse, "Adaptive Allocation of Decision Making Responsibilities Between Supervisor and Computer," Proceedings of the International Symposium on Monitoring Behavior and Supervisory Control, Berchtesgaden, Germany, March 1976.
3. W. B. Rouse, "Human Interaction with an Intelligent Computer in Multi-Task Situations," Proceedings of the Eleventh Annual Conference on Manual Control, NASA Ames Research Center, pp. 130-143, May 1975.

4. K. D. Enstrom, "Real-Time Adaptive Modeling of the Human Controller with Application to Man-Computer Interaction," MSIE Thesis, University of Illinois at Urbana-Champaign, 1976.
5. D. McRuer and D. H. Weir, "Theory of Manual Vehicular Controls," IEEE Transactions on Man-Machine Systems, Vol. MMS-10, No. 4, pp. 257-291, 1969.
6. T. B. Sheridan and W. R. Ferrell, Man Machine Systems, Cambridge, Mass.: MIT Press, 1974.
7. R. C. K. Lee, Optimal Estimation, Identification, and Control, Cambridge, Mass.: MIT Press, 1964.
8. J. M. Mendel, Discrete Techniques of Parameter Estimation, New York: Marcel Dekker, 1973.
9. N. Morrison, Introduction to Sequential Smoothing and Prediction, New York: McGraw-Hill, 1969.
10. M. M. Tatsuoka, Multivariate Analysis, New York: Wiley, 1971.

THE USE OF EVENT-RELATED-POTENTIALS IN THE ENHANCEMENT
OF SYSTEM PERFORMANCE*

By Christopher Wickens, Jack Isreal, Gregory McCarthy,
Daniel Gopher and Emanuel Donchin

Cognitive Psychophysiology Laboratory
Department of Psychology

and Aviation Research Laboratory

University of Illinois

ABSTRACT

Advancing computer technology has facilitated the implementation of on-line adaptive man-machine systems. In these systems, computer decisions based on information concerning the state of the operator can affect the nature of the man-machine interaction. Some limitations of performance measures as sources to provide information to the computer are presented, and it is argued that these limitations are particularly restrictive when the information concerns selective attention in a monitoring task, or the assessment of residual attention. It is suggested that psychophysiological measures such as the event related cortical potential (ERP) may be utilized to bypass these limitations. The characteristics of the ERP and experimental demonstrations of its sensitivity to attentional manipulation are described and a program of relevant research is then outlined. This includes a description of a pilot investigation of the relationship between the ERP to auditory probe stimuli and the workload demands of a two dimensional tracking task.

PSYCHOPHYSIOLOGICAL MEASURES IN THE ADAPTIVE MAN-MACHINE SYSTEM

The remarkable developments in mini- and micro-computers is transforming the design of man-machine systems. The computer industry is producing smaller, faster and more economical computers. It is, therefore, increasingly easier to incorporate computers as control components in man-machine systems with striking effects on the flexibility of the systems. Most notable is the increasing prevalence of adaptive systems. In these the computer can affect the nature of the man-machine interaction by implementing dynamically, an optimizing algorithm. System behavior can be adjusted to the continually changing states of the operator, the environment and to the

*This research is sponsored by the Cybernetic Technology Office of ARPA monitored by ONR thru Contract number N-000-14-76-C-0002, (Principal Investigator, E. Donchin).

interaction between the two. This flexibility depends on an exchange of information between the operator and the machine. Within the framework of a computer based system it is possible to conceive of novel channels of communication to supplement conventional display and control interfaces. This report describes a program, which is one of several Biocybernetic programs supported by ARPA's Cybernetic Technology Office, which seeks to develop such a communication channel utilizing information which can be derived from brain-waves.

The environment for which we developed these channels is exemplified by the hypothetical high performance aircraft with on-board computer facilities represented in Figure 1. The conventional Display, Pilot, Control and Plant components are supplemented by various performance aids which can be implemented or adjusted on-line. These may consist of the addition or removal of predictive display information, an alteration of the Control or Plant dynamics along various axes or perhaps the assumption of control of certain tasks normally performed by the operator. In order that these adaptive decisions be made intelligently by the on-board aiding center, a managing computer or Decision Center shown in Figure 1 must be well informed. The information it requires includes obviously characteristics of the mission, the status of other aircraft, vehicular disturbances and ground controlled inputs. Additionally, it would be extremely useful to the Decision Center to have estimates of two important aspects of the operator's attentional state: what information he is processing or ignoring at any time (selective attention), and his overall level of workload or involvement with on-going tasks. Operator workload in this sense is often conceptualized as a variable that is reciprocally related to the amount of residual attention available to deal with unexpected environmental contingencies (1).

Various performance measures are traditionally used to index attention and workload. These have been found to provide adequate indices of both aspects of the attentional state. For example, in a multi-display selective attention task, the allocation of attention between sources of discrete stimuli has been revealed by response latency (2, 3) or accuracy (4). Attention allocation to continuous tasks has been successfully identified through changes in tracking gain (2, 4), information transmitted (4), time delay (5), holds in the tracking output (6, 7), remnant or observation noise (8, 9), or by more complex coefficients of a linear discrete time series model (10).

Operator workload or residual attention has often been measured by the "secondary task loading" technique (11). The operator is presumed to possess some limited quantity of processing resources which can be distributed among various tasks. As a primary task is made more difficult--its workload demand increased--a greater quantity of the limited resource is required to maintain criterion performance, and conversely a lesser amount remains to perform the secondary task. As a consequence, secondary task performance decreases, serving as an index of primary task difficulty. Such techniques have proven reliable in comparing different display or control configurations (12), or in validating subjective estimates of control task difficulty (13).

While performance measures do serve adequately as indices of selective attention, their use is necessarily restricted to tasks in which overt

FUNCTIONAL RELATIONSHIP AMONG ELEMENTS IN THE SYSTEM

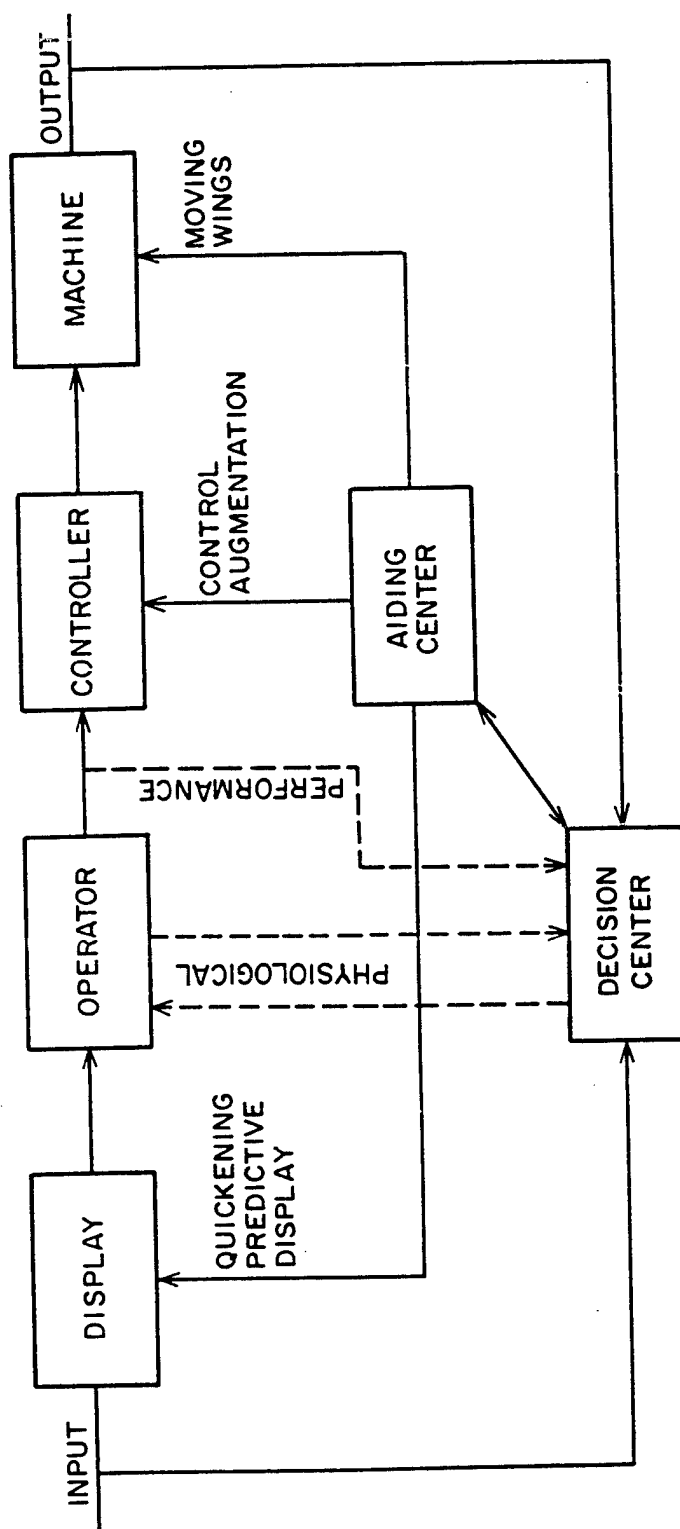


Figure 1

responses are produced. Thus they are inappropriate for a class of tasks that are becoming critically important in modern system control: passive monitoring during which few overt responses are emitted by the operator. To assess the allocation of attention during a monitoring task, periodic probe events must be inserted in the environment and an overt (detection) response to the probes required. Such probes are of necessity disruptive to ongoing performance. The difficulties encountered by an on-line assessment of workload through secondary task loading are more severe. The secondary task performance may well disrupt performance on the primary task with possibly critical consequences. It may also "saturate" the residual attention that it was designed to assess.

Even in situations where the above restrictions are not encountered, (for example divided attention between two information transmission tasks), a further limitation upon the usefulness of performance measures is presented by the inherent response variability which precludes reliable estimates of attention-sensitive parameters from a single observation. Thus assessment of any or all of the performance measures described above must involve a number of behavioral observations taken over time, a requirement which further limits the usefulness of these measures in closing an on-line adaptive loop such as that shown in Figure 1.

The shortcomings of performance measures point to an urgent need for additional sources of information which can either supplement or, as in the monitoring and workload-measurement situations described above, replace the performance measures in providing information to the Decision Center. It is for this reason that we have initiated an investigation to determine if psychophysiological measures can serve as indices of human information processing characteristics. Psychophysiological measures have two important properties that counteract the drawbacks of performance variables pointed out above. (These advantages are, of course, purchased at the cost of increased complexity in measurement. The cost effectiveness of these procedures is a matter for future research and will not be discussed here.) It is reasonable to assume that the inherent random noise sources which perturb, or contribute to the variability of psychophysiological measures are relatively uncorrelated with the sources of performance variability. Assuming an independence of the noise sources, the information provided jointly by both signals in any given observation period should be of greater reliability than the information provided by either source alone. In other words the time required for a Decision Center to obtain an estimate of the subject's internal state at a given level of reliability will be reduced when estimates are based upon joint measures.

The second advantageous feature of psychophysiological measures is that, with proper instrumentation, they may be assessed in a manner that is less disruptive to ongoing performance than the monitoring "probes" or the secondary tasks discussed above. One example of the potential usefulness of such measures has been provided by another Biocybernetics project in which Beatty (14) has demonstrated the utility of pupil diameter as a measure of the operator's cognitive processing load. Our own efforts focus on the scalp recorded event-related-potential (ERP) as a source which may provide useful information concerning the operators cognitive state.

THE EVENT-RELATED-POTENTIAL

The ERP is a transient voltage change in the brain which is elicited by any discrete event, and which may be recorded by surface electrodes attached to the scalp. Superimposed on the ongoing EEG, the ERP extends for at least 500 msec. and is characterized by a series of distinct positive and negative-going peaks with characteristic latencies following the stimulus (15, 16). The amplitude of the different peaks appears to be sensitive to physical and informational characteristics of the stimulus. The multivariate nature of the ERP provided by the separate peaks reinforces its potential value in providing considerable information to an on-board computer.

Although normally the ERP to a single stimulus is masked by the ongoing EEG voltage, rendering it difficult to identify from a visual record, multitrial averaging techniques can serve to cancel the noise contribution and provide an estimated ERP associated with an event. Alternatively, and of critical importance for on-line assessment of behavior, a wide variety of techniques are being developed and implemented as a part of this project which will enable successful identification and classification of the characteristics of an ERP on a single trial (17, 18). Such techniques include frequency domain filtering of EEG activity, iterative time domain adjustment to reveal peaks (19) or application of linear discriminant analysis (20).

Although the success of ERPs as indicants of attention in a complex cockpit-like environment has not been established, there is nevertheless strong experimental evidence that components of the waveform are sensitive to attentional manipulations. Thus for example, Donchin and Cohen (21) have found that the amplitude of the late positive peak of the waveform (P₃₀₀) reflects the allocation of attention to each of the elements in a two element visual display. This finding has been replicated in the auditory modality (22, 23). Earlier components of the waveform have also been found to be enhanced by focussed attention (24, 25). Clearly then, the evoked potential does reflect all-or-none discrete shifts in attention as defined by the relevance or non-relevance of an information source.

There is much less validation in the literature of the ability of the ERP measures to distinguish between levels of workload or attention on a continuous or graded basis in a manner which has been established with performance variables (e.g., 2, 5). In a study in which ERPs were recorded to stimuli in a two-channel signal detection task, LaFayette, Dinand and Gentil (26) were able to observe changes in the early positive and negative components of the ERP as the stimuli were made more or less relevant by instructional manipulations. In a second study, they found reliable changes in the early components of the ERP to detected tones as the workload of a secondary cognitive reasoning task was manipulated.

PRELIMINARY EXPERIMENTATION

We are currently investigating the capability of ERP measures to reflect the subject's information processing characteristics in an environment that simulates more closely the control and monitoring tasks confronted by the pilot. Our basic experimental approach is to record ERPs to probe auditory or visual stimuli. These probes are either irrelevant to the tasks performed (and may therefore be ignored by the subjects), or require only a minimum amount of cognitive processing, thus avoiding any disruption of primary task performance. In a selective attention/monitoring paradigm, the probe stimulus may occur along one of the relevant channels, displays or information sources to determine if the elicited ERP will reflect the extent to which that source is being processed. The stimulus attribute of the probe will however be different from the attributes of the channel that is relevant to the monitoring task. That is, if the subject is monitoring a visual channel for a spatially defined event, the probe will involve an intensity change. Alternatively, in a divided attention processing task, the ERP-eliciting stimuli can be the same stimuli as those that are processed and responded to in the performance of the tasks. Finally, in a workload manipulation paradigm, the probes are presented along channels that are totally irrelevant to the primary task performed, in order to determine if the elicited ERP's will reflect the residual attention available from that task as its workload is manipulated.

We have investigated ERP's and residual attention in a pilot experiment in which six subjects performed a two-dimensional pursuit tracking task with dynamics of the form

$$Y = \frac{K_1 + K_2 S}{S^2}$$

on both axes. Twelve 3-minute practice trials were first presented to bring the subjects to a stable level of performance via adaptive techniques with forcing function cutoff frequency employed as the adaptive variable.

ERP's were next recorded in two workload manipulation sessions. The probes consisted of a Bernoulli series of rare and frequent tones differing in pitch. The two sessions differed from each other according to whether the probe stimuli were ignored or processed. "Processing" involved maintaining an internal count of the number of rare stimuli that occurred during a trial (see Figure 2).

Within each session, workload was manipulated by two different methods (Figure 3). First, the forcing function bandwidth was increased and decreased by 30% from the asymptotic level, achieved by each subject in the adaptive practice session. This generated 3 different levels of objective task difficulty. Then, based upon the performance of each subject in the intermediate and high bandwidth conditions (moderate and large error respectively), two target diameters were selected, equal in value to the RMS error measured for each subject in those two conditions. A third, larger diameter was also selected of proportionately greater size. In this

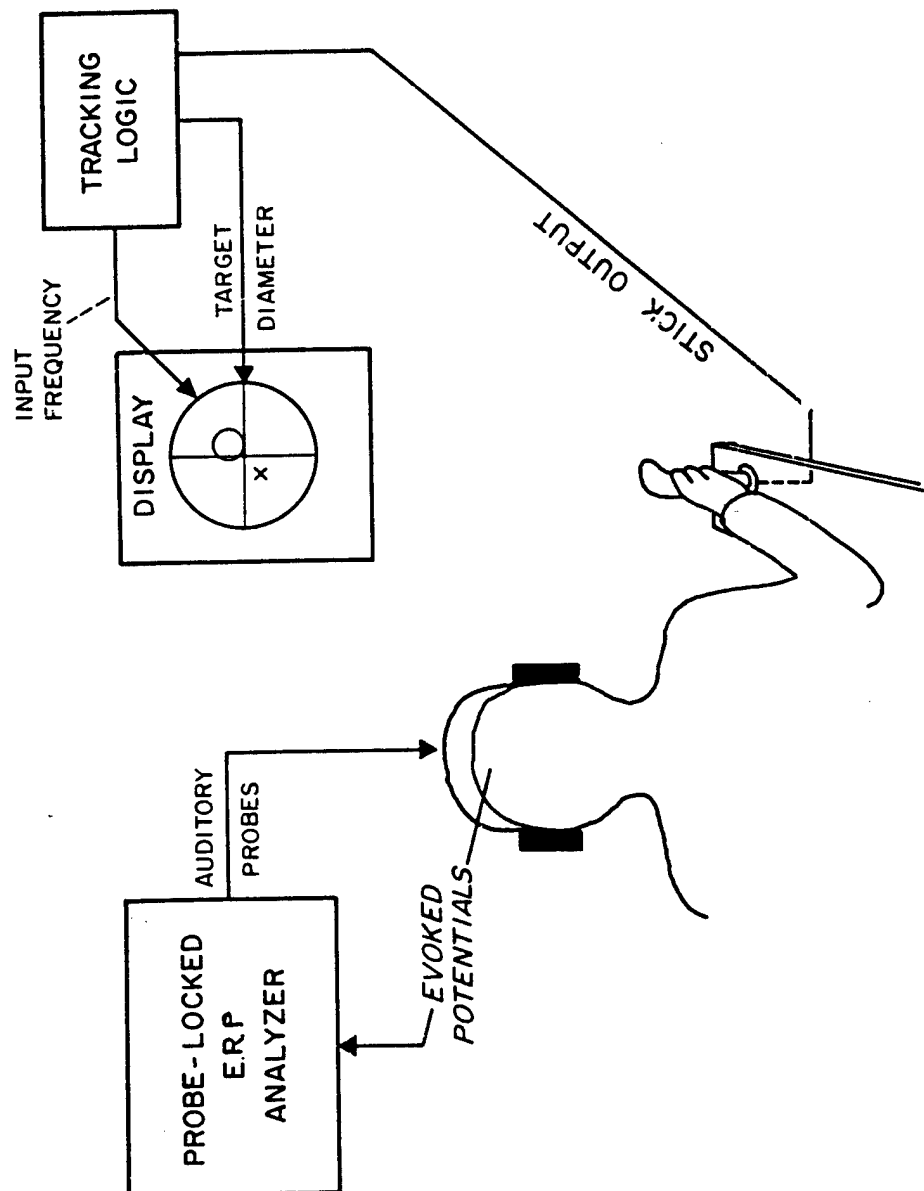


Figure 2

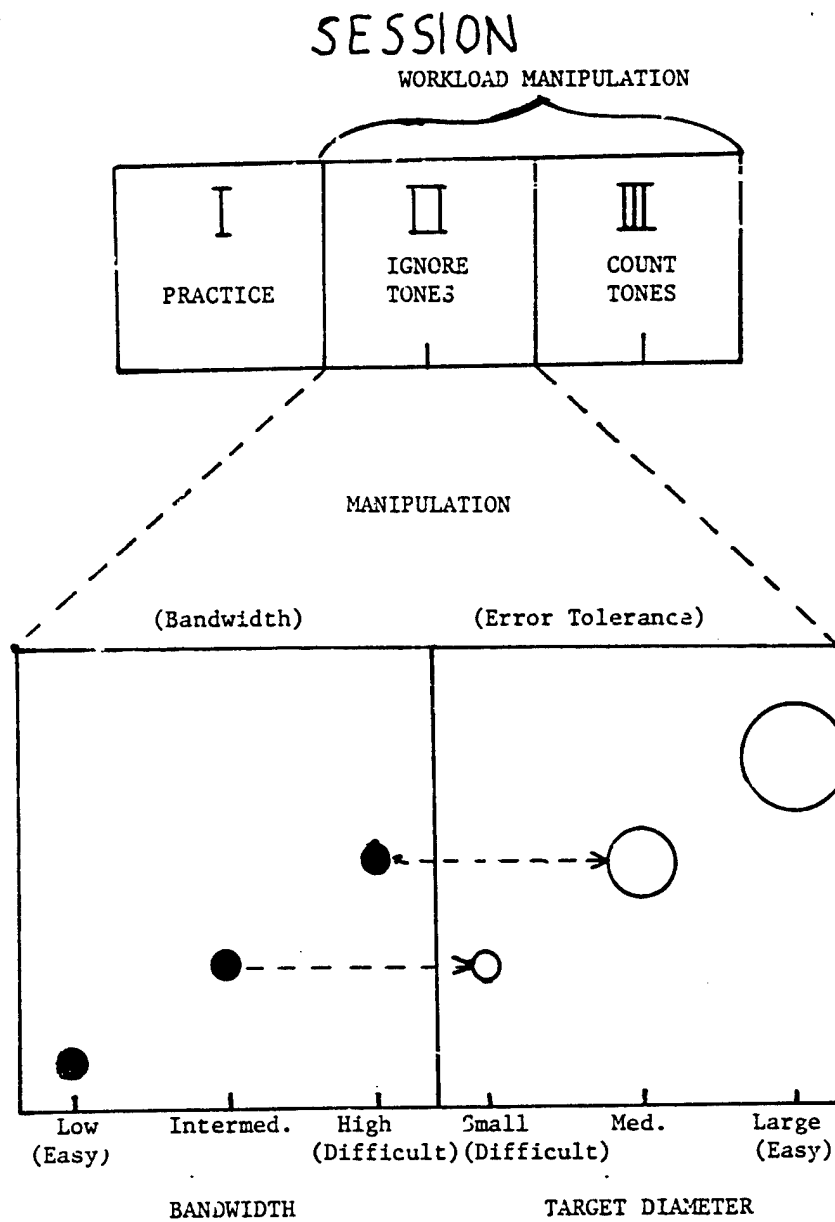


Figure 3
Experimental Design

manner the manipulation of target diameter, or required error tolerance was "calibrated" for each subject according to his sensitivity to the bandwidth manipulations, the object of this calibration being to obtain equivalent manipulations of subjective performance demands across all subjects. Tracking under the three target sizes was performed at the constant intermediate bandwidth level. It was therefore assumed that progressively more residual attention would be made available from the tracking task as the error tolerance (and obtained error) was increased.

The two particular workload manipulations employed may be placed in context by assuming that the attentional demands or subjective difficulty of a task is a joint function of its objective difficulty (e.g., task characteristics such as input bandwidth or dynamics), and the performance level required (in the present case, specified by the target diameter). What we have done then is to manipulate each of these dimensions of difficulty separately, while holding the other constant.

The logic of the tracking task, probe signal presentation and ERP recording was controlled by a PDP 11/40 computer (27). ERP's recorded from 3 electrode sites were amplified and were displayed on-line via a GT-44 graphics display terminal. The data was also recorded on digital tape for later off-line analysis and plotting by a Harris computer.

The preliminary data thus far collected has indicated that stable ERP waveforms can in fact be elicited by probes under the high levels of primary task workload required in the experiment. At this writing the data are being analyzed and it is already clear that there is a substantial degree of individual differences in the shape and temporal characteristics of the waveforms. These may well be related to different strategies that subjects adopt in performing their assigned tasks, and these strategies will be investigated through future analysis of the tracking data in the time and frequency domain. It appears however that there is some consistency within the waveforms of a given subject. In this case the between-subject variability presents no serious difficulty and may well prove useful in assessing individual differences in performance. In a sense, calibrating an ERP analyzer to the physiological response characteristics of an individual operator imposes no greater engineering demands than custom designing the helmet or flight suit to his anthropometric characteristics.

At this point, the state of our research is too preliminary to draw firm conclusions concerning the effects of attentional manipulations upon the event-related-potentials. However, given the projected importance of nondisruptive measures, and the established sensitivity of such measures to certain attentional variations in the laboratory, we are sufficiently encouraged to pursue the direction of experimentation outlined above.

REFERENCES

1. Kahneman, D. Attention and Effort. New York: Prentice Hall, 1973.
2. Gopher, D., & Wickens, C. Strategies and attention allocation in dual-task performance. Proceedings 11th Annual Conference on Manual

- Control. NASA TMX-62, 464. Washington, D.C.: U.S. Government Printing Office, 1976.
3. Ephrath, A. Detection of system failures in multi-axis tasks. Proceedings 11th Annual Conference on Manual Control. NASA TMX-62, 464. Washington, D.C.: U.S. Government Printing Office, 1976.
 4. Baty, D. L. Human transinformation rates during one-to-four axis tracking. Proceedings 7th Annual Conference on Manual Control. NASA, SP-281. Washington, D.C.: U.S. Government Printing Office, 1972.
 5. Jex, H. Two applications of a critical-instability task to secondary work load research. IEEE Transactions on Human Factors in Electronics, 1967, HFE-8, 279-282.
 6. Onstott, C. D. Task interference in multi-axis aircraft stabilization. Proceedings 12th Annual Conference on Manual Control. Urbana, Illinois, May, 1976.
 7. Cliffe, R. C. The effects of attention sharing in a dynamic dual-task environment. Proceedings 7th Annual Conference on Manual Control. NASA SP-281. Washington, D.C.: U.S. Government Printing Office, 1972.
 8. Levison, W. H., Elkind, J. I., & Ward, S. L. Studies of Multivariate Manual Control Systems: A Model for Task Interference (NASA CR-1746) Bolt Beranek & Newman, May 1971.
 9. Wickens, C. D. The effects of divided attention on information processing in manual tracking. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 1-13.
 10. Engstrom, K. D., & Rouse, W. B. Telling a computer how a human has allocated his attention between control and monitoring tasks. Proceedings 12th Annual Conference on Manual Control. Urbana, Illinois, May, 1976.
 11. Rolfe, J. M. The secondary task as a measure of mental load. In W. T. Singleton, R. S. Easterby, and D. E. Whitfield (Eds.), Measurement of Man at Work. London: Taylor & Francis, 1971.
 12. Krauss, E. F., & Roscoe, S. N. Reorganization of airplane manual control flight dynamics. In W. B. Knowles, M. S. Sanders & F. A. Muckler (Eds.), Proceedings 16th Annual Meeting of the Human Factors Society. Santa Monica, California: October, 1972.
 13. Clement, W., McRuer, D., & Klein, R. Application of New Pilot Workload Measure to Manual Control Display Design. System Technology, Inc. Paper No. 133, April 1973.
 14. Beatty, J. Pupillometric measurement of cognitive workload in complex man/machine systems. Proceedings 12th Annual Conference on Manual Control. Urbana, Illinois, May 1976.
 15. Donchin, E. Brain electrical correlates of pattern recognition. In G. F. Inbar (Ed.), Signal Analysis and Pattern Recognition in Biomedical Engineering. New York: John Wiley, 1975. Pp. 199-218.

16. Regan, D. Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. London: Chapman and Hall, Ltd., 1972.
17. Donchin, E., & Herning, R. I. A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event related potentials. Electroencephalography and Clinical Neurophysiology, 1975, 38, 51-68.
18. Squires, K. C., & Donchin, E. Beyond averaging: The use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. Electroencephalography and Clinical Neurophysiology, 1976, In press.
19. Woody, C. D. Characterization of an adaptive filter for the analysis of variable latency neuro-electric signals. Medical and Biological Engineering, 1967, 5, 539-553.
20. Donchin, E. Discriminant analysis in averaged evoked response studies: The study of single trial data. Electroencephalography and Clinical Neurophysiology, 1969, 27, 311-314.
21. Donchin, E., & Cohen, L. Average evoked potentials and intramodality selective attention. Electroencephalography and Clinical Neurophysiology, 1967, 22, 537-546.
22. Smith, D. B., Donchin, E., Cohen, L., & Starr, A. Auditory averaged evoked potentials in man during selective binaural listening. Electroencephalography and Clinical Neurophysiology, 1970, 28, 146-152.
23. Squires, K., Donchin, E., McCarthy, G., & Herning, R. On the influence of task relevance and stimulus probability on ERP components. Electroencephalography and Clinical Neurophysiology, 1976, In press.
24. Schwent, V., & Hillyard, S. A. Auditory evoked potentials during multi-channel selective listening role of pitch and localization cues. Electroencephalography and Clinical Neurophysiology, 1975, 38, 131-138.
25. Wilkinson, R. T., & Lee, M. V. Auditory evoked potentials and selective attention. Electroencephalography and Clinical Neurophysiology, 1972, 33, 411-418.
26. LaFayette, Dinand & Gentil. Average evoked potentials in relation to attitude, mental load, and intelligence. In W. T. Singleton, R. S. Easterby, and D. E. Whitfield (Eds.), Measurement of Man at Work. London: Taylor & Francis, 1971.
27. Donchin, E., & Heffley, E. Minicomputers in the signal-averaging laboratory. American Psychologist, 1975, 30, 299-312.

PUPILLOMETRIC MEASUREMENT OF COGNITIVE WORKLOAD

By Jackson Beatty

Department of Psychology
University of California at Los Angeles

SUMMARY

The momentary workload that is imposed by a cognitive task upon the limited capacity human information-processing system appears to be accurately reflected in the momentary level of central nervous system activation. The utility of pupillometric methods of workload assessment is evaluated and several lines of experimental evidence relating activation and cognitive function are reviewed.

INTRODUCTION

Information processing tasks differ in the extent and duration of the demands that they place upon the limited capacity of the human nervous system to handle information. For most tasks, processing demands are not constant, but vary from moment to moment in response to changes in the functional organization of the task. These demands may be thought to represent the cognitive workload associated with the task, a time-varying function of the demand for limited resources.

Given the assumption that cognitive capacity is fixed (reference 1), the momentary demands of any single processing function for capacity may be estimated by determining the amount of residual capacity that may be allocated to another processing task that is assigned a secondary priority (reference 2). Secondary-task measurement of cognitive workload is of major importance in the study of both cognitive capacity and the resource demands of particular processes, but both technical (reference 2) and theoretical (reference 3) difficulties preclude the utilization of secondary-task procedures in many situations. For this reason the more convenient method of subjective estimation of cognitive workload is still commonly employed (reference 4) despite serious questions as to both the reliability and validity of such rating procedures.

A third approach to the problem of measuring momentary cognitive workload stems from the observation that momentary workload is directly reflected in the mo-

mentary level of central nervous system (CNS) activation (references 5 and 6). Of the various indicators of activation, pupillometric measurement techniques (references 7, 8 and 9) appear to be most sensitive and reliable (reference 10).

The present paper examines several lines of evidence suggesting that pupillometric measures of activation serve as a reliable indicator of cognitive workload in perception, memory, decision and complex problem solving. An extension of this experimental method to the study of problems of workload optimization in complex man/machine systems is then considered.

PERCEPTUAL PROCESSES

Perceptual processes appear to proceed quite effortlessly and place rather little demand upon the limited capacity of the human information-processing system (5). Thus Wickens (reference 11) was unable to observe a secondary task decrement when a sensory signal-detection task was imposed as the primary task in an experiment investigating the distribution of processing capacity. The workload involved in the detection of weak signals is quite small.

In this context, it is of interest to note that small but reliable pupillary dilations accompany the detection of both visual and acoustic signals at near-threshold intensities. Hakerem and Sutton (reference 12) examined the pupillary movements that accompany the perception of weak visual stimuli and were able to show a dilation for signals that were detected which was absent for signals that were missed. More recently Beatty and Wagoner (reference 13) provided a pupillometric analysis of activation in the detection of weak acoustic signals using a rating-scale response procedure (see reference 14). Using unmarked observation intervals, no pupillary dilations were observed in the absence of a signal regardless of the outcome of the observer's decision. In the presence of a signal, a dilation of the pupil appeared in the interval between signal delivery and response cue onset. The magnitude of this dilation varied monotonically with the observer's rated probability that a signal had been presented.

These data raise the interesting possibility that pupillometric methods may provide a more sensitive measure of cognitive load than do conventional secondary-task measurement techniques. Thus the small pupillary dilations observed during perceptual processing may be indexing brain workload levels that are not of sufficient magnitude to be detected by secondary task interference methods.

DECISION PROCESSES

Even simple decision processes appear to impose some workload on the cognitive system as indicated by pupillometric measures of activation. For example, Simpson and Hale (reference 15) measured pupillary diameter in two groups of subjects who were required to move a lever to one of four positions. In the decision group, subjects were told at the beginning of each trial that either of two directions was permissible (e.g., front or left). Seven seconds later a response cue was presented and the subject initiated one of the two movements. In the no-decision control group, subjects were instructed exactly as to the desired movement on each trial (e.g., front). Pupillary dilation in the post-instruction pre-response period was larger and more prolonged for those subjects who had to choose between two movements before responding.

Substantially larger pupillary dilations are observed to accompany more difficult decision processes. In an experiment reported by Kahneman and Beatty (reference 16), listeners were required to determine whether a comparison tone was of higher or lower pitch than the standard. Clear pupillary dilation occurred in the 4-second decision period between presentation of the comparison tone and the response cue. The amplitude of this dilation varied as a direct function of decision difficulty, the difference in frequency between the standard (850 Hz) and comparison tones. This relation is shown in figure 1, which presents both the amplitude of dilation in the decision period and the percent decision errors as a function of the frequency of the comparison tone. These dilations were highly reliable and did not habituate over the experimental session. Pupillary dilations during decision appear to vary as a function of cognitive workload, as inferred from task parameters and performance data.

MEMORY PROCESSES

The idea that human information-processing capacity is limited arose directly from the study of the limitations of human short-term or working memory (reference 17). Our capacity for unrelated items is on the order of seven or eight, with some adjustment being made for the difficulty of the to-be-remembered units. If pupillary movements reflect CNS activation shifts as a function of cognitive workload, then these relations should be clearly revealed in the pupillometric investigation of memory processes.

Kahneman and Beatty (reference 18) provided a demonstration that the momen-

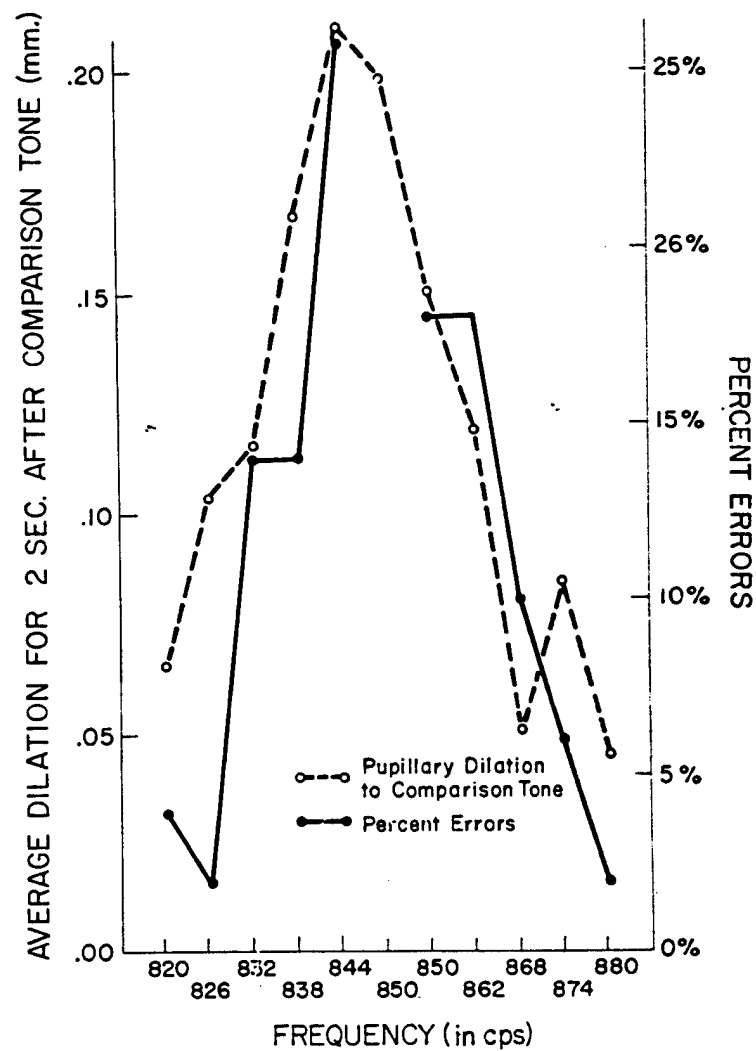


Figure 1. Average pupillary dilation during the decision period and percent errors as a function of the frequency of the comparison tone. The frequency of the standard was 850 cps. (From Kahneman & Beatty, 1967)

tary load placed upon the cognitive system by a memory task is reflected in pupillary diameter. In a series of experiments on short-term serial memory using paced recall, students were required to listen to strings of from one to seven items and, after a 2 second pause, repeat the string at the rate of one item per second. For strings of digits, pupillary diameter increased as each item of the input string was heard and decreased as each item of the output string was spoken. Thus pupillary diameter at the pause between input and output varied as a monotonic function of the number of items held in memory. These pupillary functions are shown in figure 2A.

Workload in a memory task depends not only upon the number of items to be remembered, but also upon the difficulty of each of the items themselves. Thus, as fewer unrelated words may be reliably recalled than unrelated digits, the load imposed by each word upon the cognitive system is presumed to be greater. Figure 2B presents the results of a serial memory experiment involving strings of four digits or four words. For the simple recall conditions, it is apparent that the slope of the pupillary function is greater for the more difficult word strings than for the easier digit strings. That these pupillary response functions are sensitive to processing parameters is evident from the large dilations observed under the condition labelled "transformation," in which the subject was required to respond to the string of 4 digits with another string obtained by adding 1 to each digit of the input string. This transformation task is the most difficult of all memory tasks studied, as indicated by the error data, and it consistently was accompanied by larger pupillary movements indicating CNS activation.

Behavioral data supporting the contention that the demands upon limited information-processing capacity increase during the rising phase of the pupillary response function as items are entered into working memory and decrease during the falling phase of that function as items are successively recalled from memory, is provided by an experiment in which residual capacity was measured using secondary-task measurement. Kahneman, Beatty, and Pollack (reference 19) reported that the pattern of interference with a secondary perceptual-detection task exactly paralleled the pupillary-activation curve obtained for the serial memory transformation task alone. For serial memory tasks, changes in cognitive workload appear to be reflected in the momentary level of CNS activation, as indexed by pupillometric measurement.

COMPLEX PROBLEM SOLVING

Pupillary dilations accompanying complex problem solving appear to be related directly to the difficulty of such processing, although behavioral assessments of workload have not yet appeared for these types of cognitive tasks. For example, Hess

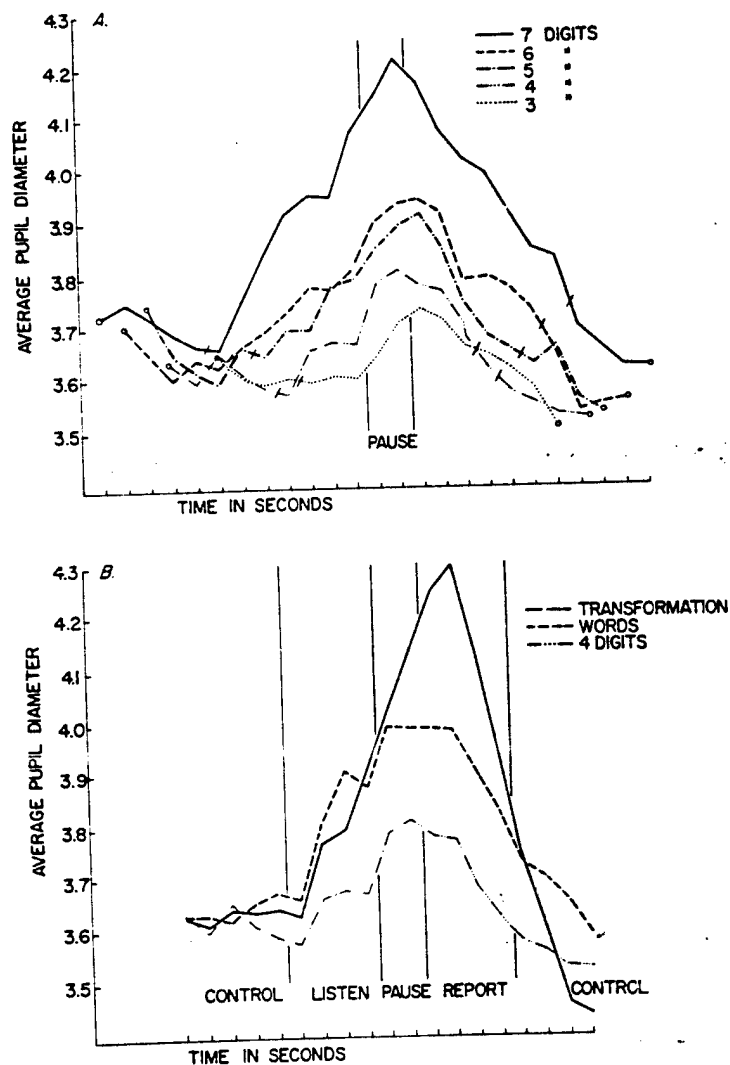


Figure 2. Upper graph: Average pupillary diameter during presentation and recall of strings of 3 to 7 digits, superimposed about the two second pause between presentation and recall. Slashes indicate the beginning and the end of the memory task. Lower graph: Pupillary diameter during presentation and recall of four digits, words and a digit transformation task. (From Kahneman & Beatty, 1966)

and Polt (8) examined pupillary movements as multiplication problems were solved mentally. Pupillary diameter increased during the period preceding solution, and the magnitude of dilation was related to presumed problem difficulty. Payne, Parry, and Harasymiw (reference 20) also report a monotonic relation between mean pupillary diameter and problem difficulty, but note that this relationship is markedly nonlinear with respect to difficulty scales based upon percent correct solution, time to solution or subjective rating of difficulty. Pupillary diameter in mental multiplication appears to peak rapidly as a function of difficulty, with more difficult problems requiring more time until solution is reached. This suggests that cognitive capacity is quite fully taxed in complex mental arithmetic problems so that the workload per unit time remains relatively constant as problem difficulty is increased over moderate levels, but that the total time to solution is increased.

Other types of complex problem solving tasks show similar relationships between pupillary dilation and problem difficulty. For example, Bradshaw (reference 21) has reported that larger pupillary dilations accompany the solving of more difficult anagrams, and that these dilations are maintained until solution is reached.

IMPLICATIONS FOR WORKLOAD EVALUATION IN MAN/MACHINE SYSTEMS

Traditional interference and subjective-rating methods of workload evaluation have been employed in the design of complex man/machine interfaces, but neither is without its own particular limitations. Pupillometric methods of workload estimation provide a third alternative that in certain situations might be preferable to either of the more traditional measurements.

One problem for which pupillometric assessment procedures appear to be well-suited is that of display evaluation. Pupillometric methods permit reliable measurement of the small cognitive workloads associated with the processing of sensory information that may not be detectable by interference methods. One project underway in our laboratories examines the effects of display readability on the pupillary dilations accompanying information acquisition. A second experiment is concerned with pupillometrically measuring cognitive workload involved in processing computer-generated speech at various levels of intelligibility.

The most intriguing possibility is that the measurement of central nervous system activation associated with cognitive function might provide a common metric for the comparison of workload in tasks that differ substantially in their functional characteristics. Underlying this possibility is the idea that CNS activation is the limited general resource that is allocated among cognitive processes demanding capacity. If this is the case, then it may be possible to directly compare perceptual, memory,

symbol manipulation and response processes in terms of activation requirements. At present, however, we may only conclude that the pupillometric measures of activation are useful in measuring cognitive load for a range of cognitive processes. No evidence concerning the comparability of measurements made across diverse processes has yet appeared.

REFERENCES

1. Broadbent, D. E.: Perception and Communication. London: Pergamon Press, 1958.
2. Kerr, B.: Processing Demands During Mental Operations. Memory & Cognition, 1973, 1, 401-412.
3. Norman, D. A.; and Bobrow, D. G.: On Data-limited and Resource-limited Processes. Cognitive Psychology, 1975, 7, 44-64.
4. McCormick, E. J.: Human Factors Engineering. (3rd ed.). New York: McGraw-Hill, 1970.
5. Kahneman, D.: Attention and Effort. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.
6. Pribram, K. H.; and McGuinness, D.: Arousal, Activation, and Effort in the Control of Attention. Psychological Review, 1975, 82, 116-149.
7. Loewenfeld, I. E.: Mechanisms of Reflex Dilations of the Pupil, Historical and Experimental Analysis. Documenta Ophthalmologica, 1958, 12, 185-448.
8. Hess, E. H.; and Polt, J. H.: Pupil Size in Relation to Mental Activity During Simple Problem Solving. Science, 1964, 143, 1190-1192.
9. Goldwater, B. C.: Psychological Significance of Pupillary Movements. Psychological Bulletin, 1972, 77, 340-355.
10. Kahneman, D.; Tursky, B.; Shapiro, D.; and Crider, A.: Pupillary, Heart Rate and Skin Resistance Changes During a Mental Task. Journal of Experimental Psychology, 1969, 79, 164-167.
11. Wickens, C. D.: The Effect of Time Sharing on the Performance of Information Processing Tasks: A Feedback Control Analysis. Technical Report No. 51. Human Performance Center, The University of Michigan, 1974.
12. Hakerem, G.; and Sutton, S.: Pupillary Response at Visual Threshold. Nature, 1966, 212, 485.
13. Beatty, J.; and Wagoner, B. L.: Pupillary Measurement of Sensory and Decision Processes in an Acoustic Signal-detection Task. Paper presented at the meeting of the Psychonomic Society, Denver, Colorado, 1975.
14. Green, D. M.; and Swets, J. A.: Signal Detection Theory and Psychophysics. New York: Wiley, 1966.
15. Simpson, H. M.; and Hale, S. M.: Pupillary Changes During a Decision-Making Task. Perceptual and Motor Skills, 1969, 29, 495-498.
16. Kahneman, D.; and Beatty, J.: Pupillary Responses in a Pitch Discrimination Task. Perception and Psychophysics, 1967, 2, 101-105.

17. Miller, G. A.: The Magical Number Seven, Plus or Minus Two. Psychological Review, 1956, 63, 81-97.
18. Kahneman, D.; and Beatty, J.: Pupil Diameter and Load on Memory. Science, 1966, 154, 1582-1585.
19. Kahneman, D.; Beatty, J.; and Pollack, E.: Perceptual Deficit During a Mental Task. Science, 1967, 157, 218-219.
20. Payne, D. T.; Parry, M. E.; and Harasymiw, S. J.: Percentage of Pupillary Dilation as a Measure of Item Difficulty. Perception and Psychophysics, 1968, 4, 139-143.
21. Bradshaw, J. L.: Pupil Size and Problem Solving. Quarterly Journal of Experimental Psychology, 1968, 20, 116-122.

SECTION III
DISPLAYS AND CONTROLS

Chairman: R. W. Allen

PRECEDING PAGE BLANK NOT FILMED

LEFT-BRAIN/RIGHT-BRAIN AND SYMBOLIC/ANALOGIC

HUMAN OPERATOR OUTPUT COMPATIBILITY

By William L. Verplank

Massachusetts Institute of Technology

SUMMARY

This paper introduces the terms symbolic (e.g., buttons, keyboards) and analogic (e.g., joysticks, light pens) to describe the range of devices for interaction between human operator (h.o.) and computer. With increasing automation, there is a tendency for man-machine communication to become increasingly symbolic. There is a danger that the interface will constrain both the communication and the thinking of the human operator. To remain natural and flexible most systems should allow for simultaneous access over a range from symbolic to analogic. This duality is suggestive of long-recognized duality in human thought and some recent evidence from split-brain patients.

COMPUTER-AIDED MANIPULATION

The structure and capability of any h.o.-computer system is strongly constrained by the form of the h.o.-machine interface. I considered that interface as the first concern in designing a system for supervisory control of remote manipulation [1]. The supervisory control concept was first suggested by Ferrell [2], and Ferrell and Sheridan [3], as a means for overcoming the difficulty of remote manipulation with transmission delay. (See figure 1.)

The Subroutine Approach

The particular form of supervisory control considered was that of providing the manipulator with the capability of an increasing repertoire of subroutines or subtasks. For example, John Hill's "ARM" program at Stanford Research Institute [4] has a "GROPE" subroutine that lowers the jaw around a block using a touch sensor on each finger (figure 2). This subroutine would be part of a larger task, such as collecting blocks. Some parts of the task would be under human- others under computer-control. These can be organized in a task hierarchy as shown in figure 3. Barber [5] wrote a special language for supervisory control. One of the difficulties, however, was that commands could only be given through the teletype. Programming simple motions was time-consuming and laborious.

This is an informal paper presented at:
12th Annual Conference on Manual Control
University of Illinois, 25-27 May, 1976.

Command Hardware

If the h.o. is to communicate arm positions directly, the most appropriate form of command might be "master-slave"-like control with a brace or replica. On the other hand, there may be occasion for the h.o. to indicate a direction, and the computer to modify the path, to avoid obstacles or simply to stop when it touches something. This seems the most elementary form of supervisory control, and would be quite useful with degraded or delayed feedback from a remote manipulator. The brace seems less appropriate at this level, for one thing, because there can no longer be always a one-to-one correspondence between the master and slave which requires re-setting the correspondence every time control is transferred from computer to h.o. At this level, some form of joystick providing rate-control seems more appropriate. G. Starr [6] is doing a detailed study of rate control; supervisory control is a potential application.

At a higher level, any frequently used subroutine could be called by the h.o. pushing a button or flipping a switch (e.g., "STOP IF YOU TOUCH", or "GROPE"). Strings of subroutines, or new programs would be more appropriately input through a special purpose or standard keyboard. Voice seems an appropriate command mode also, especially because it can be combined with others; for example, "SLOW" might lower the gains on a rate control; "EASY" might adjust the servo to produce compliance; "CAREFUL" might increase the sensitivity of touch or proximity sensors.

Symbolic and Analogic

One way to describe this range of command modes is as running from analogic (brace, joystick) to symbolic (button, keyboard, voice) along a continuum of abstraction. It seems important to provide a range of command modes in order that the system be accessible at different levels, possibly at two levels simultaneously. For example: "move in this direction" (ANALOGIC), "but stop if you touch something" (SYMBOLIC).

This division of control or command hardware into symbolic and analogic suggests the parallel between computer-aided manipulation (C.A.M.) and computer-aided design (C.A.D.) (figure 4). In both cases it seems most appropriate to use both symbolic and analogic controls perhaps simultaneously.

VISUAL THINKING

My teaching experience in the Design Division at Stanford University has made me more aware of the importance of this distinction between symbolic and analogic modes and their relationship to thinking. Robert McKim [7] has developed a drawing and design course that focuses on rapid visualization (free-hand sketching) and the process of design. The philosophy of the course (and what makes it so popular) is that productive thinking of all sorts requires fluency in not only analytical, verbal modes (on which most education is

focused) but also in the sensory, visual modes (in which our education is severely lacking).

What is visual thinking? To give you an experience in what visual thinking is, here is a problem to solve from McKim's book:

"Painted Cube: Shut your eyes. Think of a wooden cube such as a child's block. It is painted. Now imagine that you take two parallel and vertical cuts through the cube, dividing it into equal thirds. Now take two additional vertical cuts, at 90° to the first ones, dividing the cube into equal ninths. Finally, take two parallel and horizontal cuts through the cube, dividing it into 27 cubes. Now, how many of these small cubes are painted on three sides? On two sides? On one side? How many cubes are unpainted?"

Most people report quite clear use of some form of visual imagery in counting the cubes. A symbolic, mathematical approach would be cumbersome and inappropriate. One of the unique features of the Visual Thinking course is "experiential exercises for the mind's eye". Mental manipulation of mind's-eye imagery improves with practice.

Two Modes of Thinking

There has long been a distinction made between ways of thinking and great lists can be made of the dichotomies that seem to correlate. My own list is given in figure 5. Of importance to problem solving, what has been referred to as "creative" thinking is the ability to move freely from one mode to the other. The "creative insight" often comes through visual or sensory imagery. There are numerous accounts of this process, the most recent is by T.H. Krueger [8].

Split-Brain Studies

Recent clinical evidence on split-brain human patients (with hemispheres separated to relieve epilepsy) has revived interest in the differing roles of the right and left sides of the brain [9,10,11]. The left hemisphere (which controls the right hand) is usually dominant and controls speech. The right brain is associated more with sensory and spatial thinking. Whether or not the dichotomy strictly holds (both hemispheres have some language ability; vision occurs on both sides), the notion of a dual brain is a powerful metaphor for reminding us that there are distinctly different modes of knowing and thinking.

ANALOGIC PART PROGRAMMING

David Gossard, at M.I.T., has recently applied this notion of the

symbolic and analogic modes to preparing programs for numerically controlled machine tools [12]. Conventionally, a highly trained programmer prepares a program in a symbolic language (e.g., APT) directly from the drawings of the part (figure 6a). This program is compiled to produce an ideal tool-path (cutter location or C/L data) which must be verified. A post-processor computes the commands necessary to making the particular machine tool follow that path. Gossard's system (figure 6b) uses computer graphics with a picture of the work-piece, tool and clamps. The programmer uses knobs (analogic) and keyboard (symbolic) to move the tool through the appropriate paths. The computer can record and play back, straighten out, or interpret symbolically (e.g., circle) paths put in analogically by the operator. The computer then outputs the C/L data directly.

Gossard's experience with both part programmers and machinists untrained in part-programming points up the advantages of using computer inputs compatible with thinking modes. In three hours the machinists were able to produce part programs equivalent to what they would have needed three months' training to write. Even the trained programmers preferred the analogic mode. One of the unexpected advantages commented on by the users was the ability to instantly see and to avoid problems of interference between tool and clamps. The programming of complicated spatial motions and relationships is a process that clearly benefits from the ability to use both analogic and symbolic command modes.

CONCLUSION

With increasing automation, there is a tendency for man-machine communication to become increasingly symbolic. This is natural and efficient since complete processes can be commanded by simple names or labels. There is, however, the danger that the interface will constrain the thinking of the human operator to symbolic modes. The human operator also has special abilities in visual-spatial thinking. To allow both modes of thinking, the man-machine interface should provide for simultaneous access over a range of control from symbolic to analogic.

REFERENCES

1. Verplank, W.L.: Symbolic and Analogic Command Hardware for Computer-Aided Manipulation. M.S. Thesis, Massachusetts Institute of Technology, 1967.
2. Ferrell, W.R.: Remote Manipulation with Transmission Delay. Ph.D. Thesis, Massachusetts Institute of Technology, 1964.
3. Ferrell, W.R. and Sheridan, T.B.: Supervisory Control of Remote Manipulation, NEREM Record, IEEE Catalogue No. F-70, 1966 (also in: IEEE Spectrum, Vol. 4, No. 10, pp.81-88, October 1967).

4. Hill, J.W. and Sword, A.J.: "Study to Design and Develop Improved Remote Manipulator Systems, NASA Contractor Report CR 2238, Stanford Research Institute, April 1973.
5. Barber, D.J.: MANTRAN: A Symbolic Language for Supervisory Control of an Intelligent Remote Manipulator. M.S. Thesis, Massachusetts Institute of Technology, 1966.
6. Starr, G.: A Comparison of Master-Slave and Rate Control in the Presence of a Transmission Time Delay. 12th Annual Conference on Manual Control, 1976.
7. McKim, R.H.: Experiences in Visual Thinking. Belmont, California, Wadsworth Publishing Co., 1972.
8. Krueger, T.H.: Visual Imagery in Problem Solving and Scientific Creativity. Connecticut, Seal Press, 1976.
9. Gazziniga, M.S.: The Split Brain in Man, Scientific American, Vol. 217, No. 2, pp.24-29 (August 1967) (reprinted in [11]).
10. Bogen, J.E.: The Other Side of the Brain: An Appositional Mind, Bulletin of the Los Angeles Neurological Societies, Vol. 34, No. 3, pp.135-167, July 1969 (reprinted in [11]).
11. Ornstein, R.E. (Ed.): The Nature of Consciousness. San Francisco, W.H. Freeman and Co., 1973.
12. Gossard, D.C.: Analogic Part Programming with Interactive Graphics. Ph.D. Thesis, Massachusetts Institute of Technology, 1975.

SUPERVISORY CONTROL

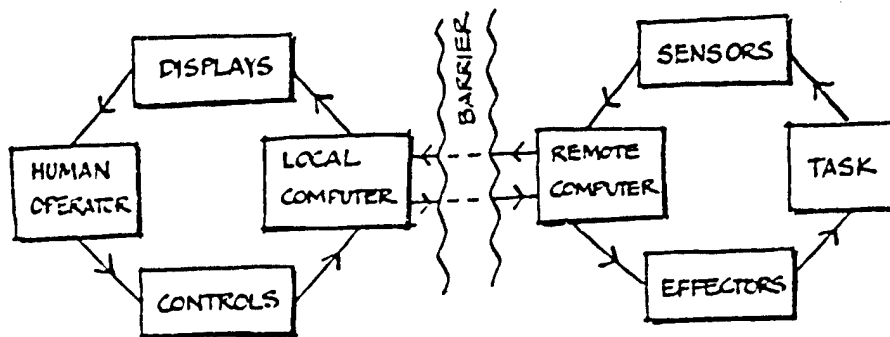
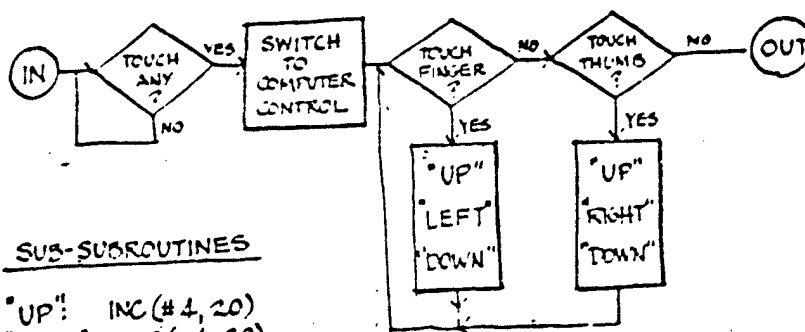


Figure 1. The supervisory control concept for remote manipulation was introduced by Ferrell and Sheridan [2,3].

SUBROUTINE "GROPE"



SUB-SUBROUTINES

"UP": INC(#4,20)
 "DOWN": DEC(#4,20)
 "LEFT": INC(#1,10)
 "RIGHT": DEC(#1,10)

Figure 2. Subroutine "GROPE" uses touch sensors on finger and thumb to center and lower the jaw around an object.

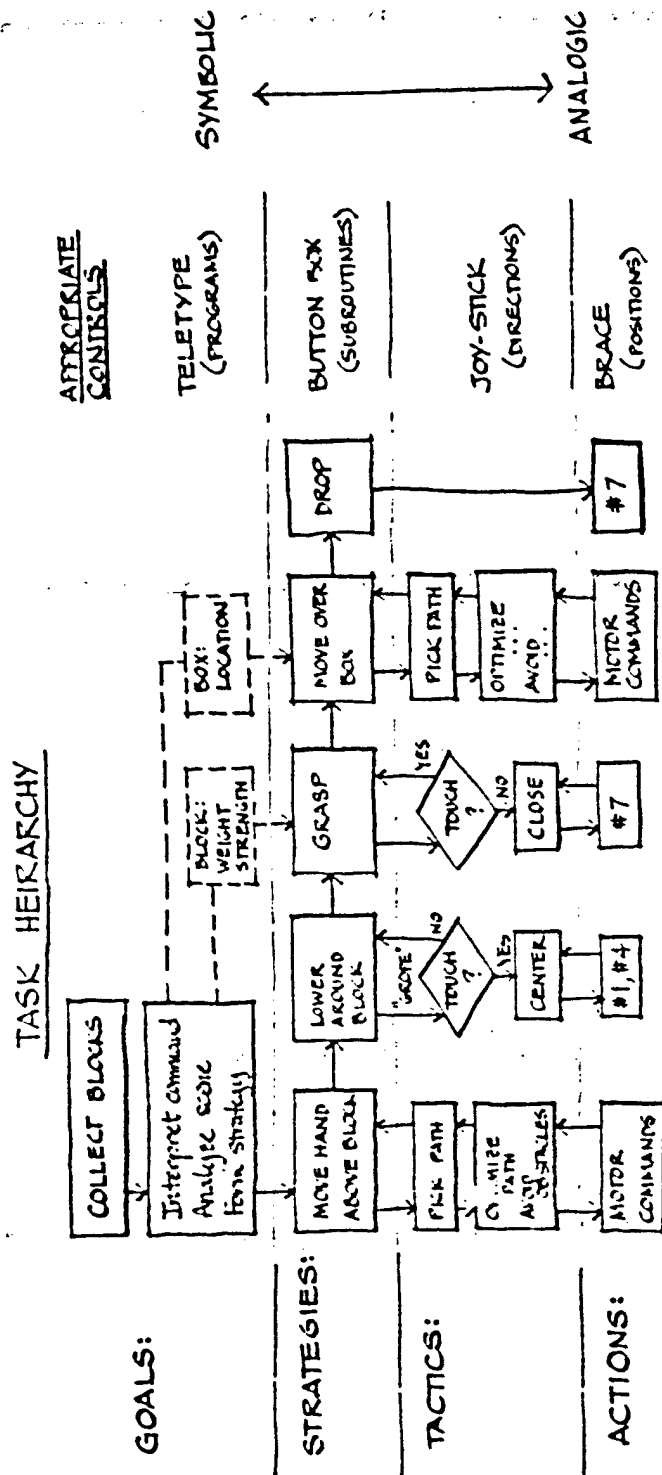
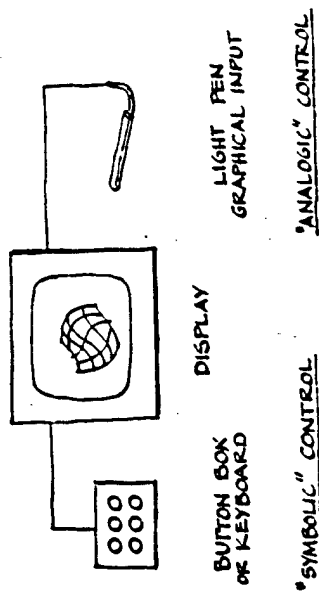
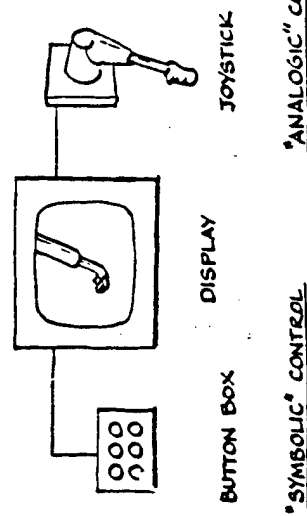


Figure 3. A task hierarchy for picking up blocks shows the interaction of h.o. and computer on a variety of levels. The command hardware appropriate at different levels is shown on the right.



4a. Command Structure for Computer-aided Design (C.A.D.)



4b. Proposed Command Hardware for Computer-aided Manipulation (C.A.M.)

Figure 4. Comparison of C.A.D. and C.A.M.

RIGHT BRAIN	LEFT BRAIN
ANALOGIC	SYMBOLIC
VISUAL	VERBAL
CONCRETE	ABSTRACT
SYNTHETIC	ANALYTICAL
ARTISTIC	SCIENTIFIC
GEOMETRICAL	LOGICAL
UNCONSCIOUS	CONSCIOUS
INTUITIVE	RATIONAL
METAPHORICAL	LITERAL
APPOSITIONAL	PROPOSITIONAL
SINISTER	ADROIT
LEFT HAND	RIGHT HAND

Figure 5. The dichotomies of thinking modes are associated with each other and the hemispheres of the brain.

Figure 6a. Conventional part programming for numerically controlled (N/C) machine tools requires a part programmer trained in a highly symbolic computer language.

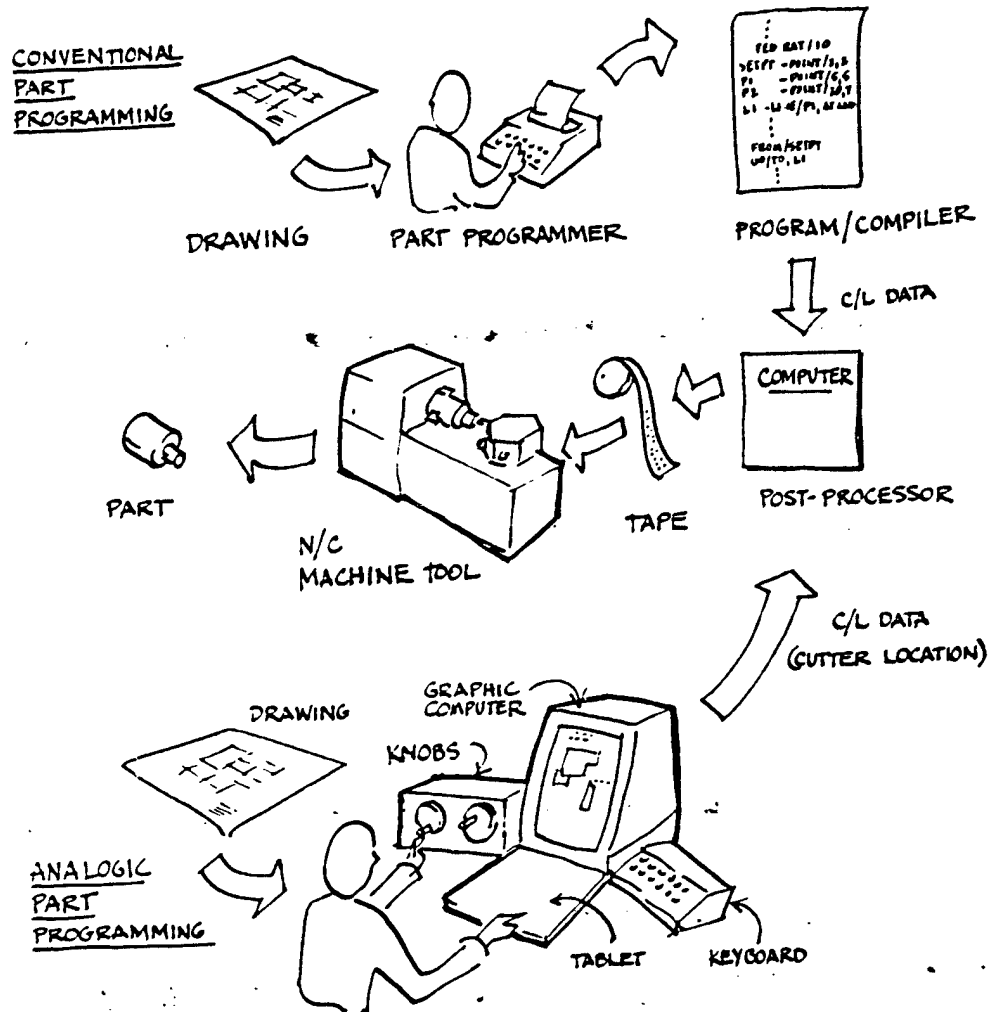


Figure 6b. Analogic Part Programming [12] uses computer displays and analogic (knobs) and symbolic (keyboard) modes for more efficient programming (less training, better programs). There appears to be better compatibility between thinking modes and human operator output (right-brain vs. left-brain and spatial vs. linguistic and analogic vs. symbolic).

A REGRESSION APPROACH TO GENERATE AIRCRAFT PREDICTOR INFORMATION

By Paul D. Gallaher, Roger A. Hunt, and Robert C. Williges

University of Illinois at Urbana-Champaign

SUMMARY

A predictor display shows the human operator future consequences of his immediate control inputs. A contact analog aircraft display is described in which an airplane-like predictor symbol depicts future airplane position and orientation. The standard method for obtaining the predictor information is to use a complete, fast-time model of the controlled vehicle. An alternative approach is presented in this paper in which least-squares, first-order, linear approximations for each of the six degrees of freedom of aircraft motion were calculated. Thirteen variables representing changes in positions and rate of change of positions were selected as parameters for the prediction equations. Separate sets of equations were determined for 7, 14, and 21 seconds prediction times and continuous, 1, and 3 seconds control neutralization assumption times. The advantages and disadvantages of this regression approach are discussed.

INTRODUCTION

Predictor displays provide the operator of manually controlled systems with information about the future state of the variable being controlled. Often this information can be generated by an analog of the system to be controlled, operating repetitively in an accelerated time scale. Ideally, to generate a predictor model using such a fast-time model, the model should be a duplicate of the original plant. For example, to put a predictor display in an aircraft trainer which uses an analog computer for all flight equations and dynamics, a second analog computer just like the first with speeded-up time constants could be used. Such complexity in using an accurate fast-time model imposes a penalty of cost, computer weight, and power requirements. In fact, Kelly (1) pointed out that it may not be necessary to have the complete accuracy of a fast-time model.

Bernotat (2) used a Taylor series expansion rather than the fast-time model approach, and found that even inaccurate predictions gave improved performance over no prediction in the control of a third-order undamped system following a step input. Kelley (3) found the same effect, but he also found that the useful prediction span decreased with model accuracy while learning times for effective manual control were increased. A comprehensive study of simplified models for an automatic predictive control system for aircraft landing in two dimensional sideways looking displays was conducted by Chestnut, Sollecito, and Troutman (4). They pointed out that the model may

be of either the analog or digital form, but they felt the digital approach offers more accuracy and flexibility. They also indicated that the time constants and gains of the model can be in error by two to one without excessive loss in performance.

The main effect of an inaccurate model is closely related to the predictor span. The magnitude of the errors in an inaccurate predictor can be determined analytically or experimentally if the plant can be observed directly or simulated accurately for comparison with a less accurate fast-time model. Errors farther into the future are usually compensated for by the fact that accuracy requirements on short predictions usually are greater than for long predictions. Predictor displays can also overcome the problem of accuracy when they are continuously updated. If updating is inaccurate or infrequent, the fast-time model must be that much more accurate.

This paper presents a least-squares, regression approach for determining first-order, linear approximations of accurate fast-time models used in predictor displays. Such a procedure would eliminate the need for an operational fast-time model while still providing a great deal of predictive accuracy. The accuracy of this regression approach for generating these predictor symbols is evaluated both at various prediction times and at various control input durations.

METHOD

Task

For the purpose of demonstrating the use of a regression approach to generate predictor information, an application incorporating predictor information in an aircraft system during an approach to landing was used. Because of the complexity and sluggishness of the aircraft system in the landing phase, manual performance depends heavily on the anticipatory abilities of the pilot. Under such circumstances, predictive displays might be very useful. Smith, Pence, Queen, and Wulfeck (5) demonstrated that the predictor display did improve performance in an approach to landing on an aircraft carrier. It even facilitated learning to such an extent that mean performance on transfer trials using a predictor was considerably higher than that of a control condition without the predictor.

The specific approach to landing task in this study was generated for a Singer-Link General Aviation Trainer (GAT-2) which simulates general, light, twin-engine aircraft. The predictor symbology was incorporated into a versatile computer-generated dynamic flight display developed by Artwick (6) and was part of an integrated, vertical situation display stylized in Figure 1. In addition to the situational information of runway outline, centerline, touchdown zone, and grid-line ground texture cues, three glide-slope indicators in the form of telephone-pole-shaped symbols and three discrete, airplane-like predictor symbols are shown on the display. The

predictor symbols represent the position of the aircraft at three particular future points in time (7, 14, and 21 seconds as used in this study) given a specified control input by the pilot.

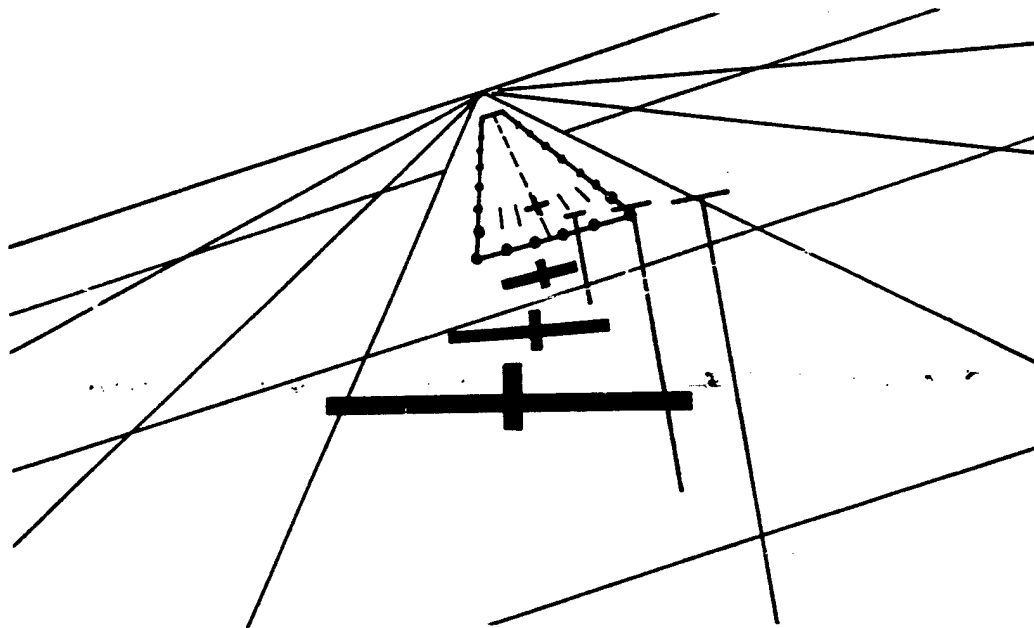


Figure 1. Stylized representation of an integrated vertical situation display showing three aircraft-like predictor symbols.

Regression Procedure

To generate the predictor symbols shown in Figure 1, one must specify the changes in the six degrees of freedom of aircraft motion as listed in Table 1. Each of these six degrees of freedom are determined by the specific flight dynamics of the aircraft. These dynamics are specified in terms of complex, higher-order differential equations which represent position, change in position, and rate of change of position as shown in Table 2. (These values are all accessible as millivolts in the GAT-2 analog computer.)

Rather than use the complete set of complex flight equations, a first-order linear approximation may suffice particularly in the limited range of variables encountered in a final approach to landing situation. A standard, least-squares, multiple linear regression analysis (Tatsuoka, 7) can be used to estimate a raw-score, linear approximation of the general form,

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m \quad (1)$$

TABLE 1

Changes in Six Degrees of Freedom of Aircraft Motion Required to Specify Aircraft Predictor Symbology

Degrees of Freedom
Change in Bank ($\Delta\theta_B$)
Change in Yaw ($\Delta\theta_Y$)
Change in Pitch ($\Delta\theta_P$)
Change in Lateral Position (Δ_X)
Change in Vertical Position (Δ_Y)
Change in Longitudinal Position (Δ_Z)

TABLE 2

Initial Variables Used to Predict Changes in Six Degrees of Freedom of Aircraft Motion

Predictor Variables
Aileron Position (α)
Rudder Position (ρ)
Elevator Position (ϵ)
Throttle Position (τ)
Bank Angle (θ_B)
Yaw Angle (θ_Y)
Pitch Angle (θ_P)
Cosine Bank ($\cos \theta_B$)
Rate of Roll ($\dot{\theta}_B$)
Rate of Pitch ($\dot{\theta}_P$)
Rate of Yaw ($\dot{\theta}_Y$)
Rate of Climb (R/C)
Velocity (v)

where Y represents the dependent variable, X_1 through X_m the independent variables, and β_0 through β_m the partial regression coefficients. Specifically, the general form of the Equation 1 for the predictor symbology case is,

$$\begin{aligned} \Delta \text{ degree of freedom} = & \beta_0 + \beta_1 \alpha + \beta_2 \rho + \beta_3 \epsilon + \beta_4 \theta_B \\ & + \beta_5 \dot{\theta}_Y + \beta_6 \dot{\theta}_P + \beta_7 \cos \theta_B + \beta_8 \dot{\theta}_B \\ & + \beta_9 \dot{\theta}_P + \beta_{10} \dot{\theta}_Y + \beta_{11} R/C \end{aligned} \quad (2)$$

where Y is replaced by the particular change in degree of freedom of interest the X's are replaced by selected variables in Table 2, and the β 's represent the raw-score, partial regression weights which are empirically determined.

All the independent variables except velocity and throttle can take on both positive and negative values. Velocity and throttle are always zero or some positive value, so their contribution to the predictor equation would always be positive. Furthermore, velocity and throttle changes should amplify the effects produced by control surface position and airplane position changes. Consequently, the independent variables of aileron, rudder, and elevator position as well as the current bank, yaw, and pitch angles shown in Equation 2 are multiplied by the velocity and throttle values of the GAT-2. The remaining four variables in Equation 2 already contain velocity and throttle information, because they are rates of change in position.

Data Collection Procedure

Of the thirteen independent variables shown in Table 2, only the changes in position of the three control surfaces (rudder, aileron, and elevator position) and the throttle position can be directly affected by the pilot. The remaining nine variables are non-linear, interacting functions of these as well as outside disturbances. For each of the four variables under direct pilot control, three levels of change in millivolts (zero, one positive, and one negative) were directly manipulated by the experimenter to obtain the necessary data for generating the regression equations. A one-third replication of the factorial combination of these four variables was observed twice resulting in 54 data collection flights. The remaining nine variables were considered to be approximately random and were not manipulated through experimenter control.

During each of these 54 data collection cycles the GAT-2 was flown in an approach to landing configuration. The landing gear was down and the proper airspeed, flap setting, manifold pressure, etc. was maintained. When the GAT-2 was flown by the pilot to the proper landing configuration, the Raytheon 704 computer maintained the control surfaces at the appropriate level, recorded the initial values of all thirteen independent variables shown in Table 2, and measured the changes in the six degrees of freedom of motion (dependent variables) after 7, 14, and 21 seconds. These latter values provided the three prediction times represented by the successive discrete predictor symbols shown in Figure 1.

To simulate the four designated control surface positions over different flights, the Ravi 704 computer was used. The analog signals from the GAT-2 representing control surface positions were intercepted prior to their use in the GAT-2 analog computer flight equations. An analog-to-digital converter made these signals available in the form of a 12-bit word. Thus, 0 to 10 volts was converted to 0 to 2048 binary. The software routines then generated changes in these signals as dictated by the one-third replicate of the factorial design. These new signals were sent through the digital-to-analog converter and into the GAT-2 computer to maintain precisely a given set of control movements.

As shown in Table 3, the factorial design of this study also allowed for the calculation of six prediction equations for each of three control assumption times at the 7, 14, and 21 second prediction times. The length of time these control surface changes were maintained prior to neutralization determined the control assumption times. When the control changes were maintained continuously over the 21 second prediction span, this produced the continuous or on-line predictor model (Warner 8). If the control changes were not maintained throughout the data collection phase, an off-line predictor model is used. Two off-line models using control assumption times of 1 and 3 seconds were also investigated in this study. A different set of 54 approaches to landing were required for each control assumption time. Consequently, a total of 162 approaches were measured.

TABLE 3

Factorial Design of Control Assumptions and Prediction Times Used to Generate the Six Regression Equations Predicting the Degrees of Freedom of Motion of the Predictor Symbol

Control Assumptions (Seconds)	Prediction Times (Seconds)		
	7	14	21
Continuous (21)	(6 Regression Equations)	(6 Regression Equations)	(6 Regression Equations)
1	(6 Regression Equations)	(6 Regression Equations)	(6 Regression Equations)
3	(6 Regression Equations)	(6 Regression Equations)	(6 Regression Equations)

RESULTS

A multiple, linear regression analysis was conducted on all 11 independent variables shown in Equation 2 for each dependent variable to determine the appropriate partial-regression coefficient values. Table 3 shows that there were six equations for each predictor time and the associated control assumptions. These six equations determined the changes in the six degrees of freedom of motion for a particular predictor symbol. Because each

prediction equation required a separate regression analysis, a total of 54 regression equations were solved.

For example, Table 4 shows the general form of the six prediction equations needed to represent the airplane predictor symbol at seven seconds in the future for a three second control assumption time. Although this regression analysis was conducted on all 11 of the independent variables shown in Equation 2, only the statistically significant ($p < .05$) predictors are shown in Table 4. Similar sets of prediction equations were derived for the other treatment conditions summarized in Table 3. In each case, however, the specific set of statistically reliable partial-regression weights varied somewhat.

TABLE 4

Prediction Equations with Significant ($p < .05$) Independent Variables Used to Determine Changes in the Six Degrees of Freedom of Aircraft Motion for Seven-Second Prediction Span and Three-Second Neutralization Assumption.

$$\Delta \theta_B = \beta_0 + \beta_1 \alpha + \beta_2 \rho + \beta_3 \theta_B + \beta_4 \dot{\theta}_B + \beta_5 \dot{\theta}_Y + \beta_6 \dot{\theta}_P$$

$$\Delta \theta_Y = \beta_0 + \beta_1 \alpha + \beta_2 \theta_B + \beta_3 \dot{\theta}_B + \beta_4 \dot{\theta}_Y + \beta_5 \dot{\theta}_P$$

$$\Delta \theta_P = \beta_0 + \beta_1 \epsilon + \beta_2 \theta_B + \beta_3 \dot{\theta}_Y + \beta_4 \dot{\theta}_P + \beta_5 \dot{\theta}_P + \beta_6 R/C$$

$$\Delta X = \beta_0 + \beta_1 \alpha + \beta_2 \rho + \beta_3 \theta_B + \beta_4 \dot{\theta}_Y + \beta_5 \dot{\theta}_B + \beta_6 \dot{\theta}_Y + \beta_7 \dot{\theta}_P$$

$$\Delta Y = \beta_0 + \beta_1 \alpha + \beta_2 \epsilon + \beta_3 \cos \theta_B + \beta_4 \dot{\theta}_B + \beta_5 \dot{\theta}_P + \beta_6 R/C$$

$$\Delta Z = \beta_0 + \beta_1 \alpha + \beta_2 \epsilon + \beta_3 \cos \theta_B + \beta_4 \dot{\theta}_B + \beta_5 \dot{\theta}_P + \beta_6 R/C$$

One convenient way of assessing the goodness of fit of each of these regression equations is to calculate the multiple correlation coefficient. The square of this value represents the percent of variance accounted for by the regression equation. Table 5 summarizes the multiple correlation coefficients for each of the 54 prediction equations of this study. (For example, the multiple correlation coefficients for the six prediction equations presented in Table 4 are .96, .98, .84, .94, .79, and .85, respectively.) Note that the change in altitude (Y) is the degree of freedom of aircraft motion which resulted in the lowest multiple correlation coefficients. Generally, the one-second control assumption time and the seven-second prediction time also produced regression equations with lower predictive accuracy.

TABLE 5

Multiple Correlation Coefficients for Each Predictor Equation

Prediction Times	Bank	Yaw	Pitch	X	Y	Z
Continuous (21 Second) Control Assumption						
7	.98	.95	.97	.87	.78	.77
14	.98	.98	.98	.95	.85	.94
21	.99	.98	.97	.97	.92	.96
1 Second Control Assumption						
7	.96	.96	.87	.88	.66	.66
14	.97	.95	.94	.92	.71	.61
21	.97	.91	.94	.93	.75	.64
3 Second Control Assumption						
7	.96	.98	.84	.94	.79	.85
14	.98	.98	.91	.98	.83	.92
21	.98	.98	.95	.98	.85	.92

DISCUSSION

The overall consistently high multiple correlation coefficients obtained in this study indicate that the regression approach yields very accurate prediction equations and is a viable alternative to using the complete, fast-time model. The lower multiple correlation coefficients for the one-second control assumption is probably reflective of the fact that a one-second control input is simply too brief to account for any significant movement of the GAT-2 over the prediction interval. Likewise, the lower predictive power of the 7 second prediction times as compared to 14 and 21 seconds merely shows that the GAT-2 dynamics are such that the simulator has not completed a response to the control force inputs. The longer prediction times represent a more complete simulator response.

A simplification of this approach for application to actual aircraft would be to remove the variables representing rates of change of motion which are not normally available. Undoubtedly, this simplification would

reduce the predictive accuracy of the regression equations, because rate parameters provided significant weightings in the prediction equations. From a behavioral point of view, however, these less precise equations may not affect the pilot's performance in flying the aircraft. Additional research is needed to determine the role of predictor symbol accuracy in determining operator control inputs before the allowable degree of predictor simplification can be specified.

This approach to generating predictor symbology offers the advantages of ease of implementation, low cost, and conformity to a digitally-generated display. In fact, this method may be better than an accurate, fast-time model in the sense that time lags are no longer proportional to prediction span because of increased computations being required further into the future. Furthermore, the prediction span need not be compromised by repetition rate, updating frequency, or computing power available because any discrete prediction is as easy to make as any other.

It should be remembered that the specific prediction equations of this study pertain only to the control dynamics of the GAT-2 at the three prediction times and control assumption times varied. In other words, the regression equations are always specific to the device from which the data are collected. The approach and procedure for generating these regression equations, however, are general and can be applied to generating predictor symbology for any specific device. Obviously, there probably are situations in which a multiple linear regression may not provide an adequate representation of the true underlying system dynamics. In such instances a regression approach is still appropriate, because it can be easily extended to higher-order, polynomial regression representations of these complex functions.

ACKNOWLEDGMENTS

Contractual support for this research was provided by the Engineering Psychology Program, Office of Naval Research, Contract Number N00014-76-C-0081. Mr. Gerald Malecki, Assistant Director of Engineering Psychology Programs, was the technical monitor of this program.

REFERENCES

1. Kelly, C. R. Developing and testing the effectiveness of the predictor instrument. Santa Monica, Calif.: Dunlap and Associates, Inc., Technical Report 252-60-1, March 1960.
2. Bernotat, R. Das prinzip der voranzeige und seine anwendung in der flugführung. Zeitschrift für Flugwissenschaften, 1965, 13, 373-377.
3. Kelley, C.R. Further research on the predictor instrument. Santa Monica Calif.: Dunlap and Associates, Inc., Technical Report 252-60-2, December 1960.

4. Chestnut, H., Sollecito, W. E. and Troutman, P. H. Predictive control system application. Transaction of AIEEE, 1961, 55, 128-139.
5. Smith, R. L., Pence, G. G., Queen, J. E., Wulfeck, J. W. Effect of a predictor instrument on learning to land a simulated jet trainer. Inglewood, Calif.: Dunlap and Associates, Inc., August, 1974.
6. Artwick, B. A. A versatile computer-generated dynamic flight display. Savoy, Ill.: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Research Laboratory, Technical Report ARL-76-5/ONR-76-1, May 1976.
7. Tatsuoka, M. M. The general linear model. Institute for Personality and Ability Testing, Champaign, Ill., 1975.
8. Warner, J. A. A fundamental study of predictive display systems, NASA CR-1274, Technical Report, 1969.

INTRODUCTION TO A COORDINATED COCKPIT DISPLAY CONCEPT

By D. L. Baty

Ames Research Center, NASA
Moffett Field, California 94035

ABSTRACT

The aircraft instrument panel has evolved as an ever growing collection of subsystem indicators. Although human factors design has reduced the difficulty in interpreting the displayed parameters, it has not resulted in a quickly and accurately assimilated body of information. Some efforts have been made to integrate displays from the pilot's viewpoint as evidenced by electronic attitude director indicators and cathode ray tube (CRT) map displays. However, with these displays, full advantage has not yet been taken of the potential of current computer and display device technology. This paper describes an initial three-display configuration, based on a rationale developed earlier, that is designed to present flight status information. Primary emphasis is on the factors that contribute to efficient perception and integration of the pilot's total spatial situation. The display system consists of three orthogonal views containing both qualitative and quantitative information presented on three 17.7 x 17.7 cm (7 x 7 in.) beam-penetration color CRTs. Therefore, the displays will show different views of the same information. In addition, to increase the visual separation of the information elements, the colors red, green, and yellow are used to depict, respectively, the control, performance, and navigation categories of flight instrumentation. As a result the displays are coordinated in information and color; therefore, the name Coordinated Cockpit Display (CCD).

INTRODUCTION

The Man-Machine Integration Branch, Flight Management Program at NASA-Ames Research Center is committed to study and research of pilot procedures and pilot-systems interfaces that will be required for aircraft operating within the National Airspace System of the 1980-1990s (ref. 1). The study of pilot information and display requirements is an integral part of this program. These requirements are generated by many complex systems ranging in scope from the overall National Airspace System to individual on-board avionics systems. If the pilot is to cope efficiently with all relevant information, careful thought must be given to the method of its presentation. In a separate report (ref. 2) several system and perceptual problems were discussed and the rationale for an initial display set was evolved. Three general design goals resulted from that discussion:

1. Divorce the display configuration from individual subsystems.

2. Display information to the pilot in such a manner that he always uses the same displays regardless of the role he is actively taking.

3. Provide pictorial representations of the aircraft situation with quantitative information appropriately related to the picture.

Based on these concepts the first prototype system was designed as described in this report. The display system is a set of three, beam-penetration color cathode-ray tubes (CRTs). Since one of three orthogonal projections of the aircraft situation will appear on each CRT, the displays will show different views of the same information. The color feature is included primarily to obtain visual separation of information elements, but additional advantage is taken of this capability by differentiating control, performance, and navigation information on the three displays with red, green, and yellow, respectively. Therefore, the displays are coordinated in information and color, and the name Coordinated Cockpit Display (CCD) was chosen to emphasize this feature of the system. Changes in internal detail, but not in overall concept, can be expected in subsequent displays.

GENERAL FEATURES OF THE CCD

The three-display configuration described here is based on three orthogonal projections of the aircraft situation: (1) perpendicular to the pilot's forward line-of-sight; (2) parallel to the ground; and (3) perpendicular to the other two. Figure 1 illustrates the relationships.

The first display is most closely related to the pilot's view out the front window and is perpendicular to the earth. For the CCD system this is called the Vertical Situation Display (VSD) (fig. 1). Because the frame of reference moves in response to aircraft attitude the first CRT presentations of this type were called Electronic Attitude Director Indicators (EADI) and that designation has remained (ref. 3). This reference to attitude is too restrictive and the term vertical situation display is currently more descriptive of the broader function visualized for this display.

The second display represents the horizontal situation and is thus called the Horizontal Situation Display (HSD). This plane is parallel to the earth's surface and is the plane in which maps are commonly drawn.

The display that will show the pilot's situation in a plane perpendicular to the earth and parallel to the pilot's forward line of sight is called the Side Vertical Situation Display (SVSD). In the past, very little attention has been given to this view of the flight situation. In addition to other features to be described, this pictorial view will be ideal for explicitly showing the altitude situation, and should make it easy for the pilot to maintain altitude awareness.

The combination of these three displays unambiguously shows the total flight situation. Each display explicitly represents two dimensions in space

and shares one of those two dimensions with each of the other two displays. For example, the display elements to be shown on the SVSD will represent up/down and fore/aft situation information; the up/down dimension is also one of the VSD dimensions, and the fore/aft dimension is one of the HSD dimensions. Therefore, each of the three displays is capable of showing different views of the same information; for example, a waypoint in space with a line joining the aircraft with the waypoint. By constructing these different views of selected information the displays are tied together, or coordinated, in terms of information content. The display elements are also to be color coded according to three (perhaps four) classes of function that will be the same on all three displays.

As prime instruments, the three CRTs will be mounted in the center of the aircraft instrument panel with the display surfaces perpendicular to the pilot's line-of-sight. This is not ideal since the pilot will have to mentally rotate coordinates to correspond with the real world. However, the alternatives of positioning the scope faces parallel to the planes they represent, either at the instrument panel or closer to the pilot, present major practical difficulties. To keep this mental rotation task as simple as possible the three displays will be positioned as shown in figure 2. This is the relationship that results if the three planes depicted in figure 1 were folded outward as if they were three sides of a box.

Color Coding

As already mentioned in the Introduction, color will be used as part of this display system, primarily to obtain visual separation of the information elements. The usefulness of different colors to separate display elements is well demonstrated by current mechanical flight directors. Because monochrome CRTs lack color separation, they become visually cluttered by even a few elements. Shape, intensity, and line coding do little to relieve the problem. The beam-penetration CRTs to be used with the initial display system can generate three basic colors: red, green and yellow. (Other intermediate colors, such as orange, can also be generated, but red, green, and yellow are the most easily discriminable.) Rather than arbitrarily assigning a color to each display element, a search was made for some consistent color assignment scheme that would also fulfill the visual separation requirement. An instrument classification scheme used by the Air Force provides three categories to match the three basic colors.

Air Force Manual 51-37 divides flight instruments into three categories: control, performance, and navigation instruments (ref. 4). Loosely defined, the control instruments indicate first response to control inputs such as aircraft attitude and engine power; the performance instruments indicate the effects of changes in the control parameters, such as pitch changes resulting in altitude and airspeed changes; and the navigation instruments indicate aircraft position relative to ground references. These three categories can also be referred to as inner, middle, and outer loop control.

The colors red, green, and yellow have been assigned to control, performance, and navigation information, respectively. (This is probably not critical

from a perceptual standpoint.) Red was chosen for control information for three reasons: (1) pilot response to control requirements must be relatively quick and red is traditionally associated with a requirement for immediate action; (2) there are fewer elements of control information than is the case for performance and navigation so less demand to "look at red"; and (3) because red elements will probably require two beam tracings to attain the desired brightness level, assigning fewer elements to red will save computer time. The present green and yellow assignments were given because early color drawings of potential displays were aesthetically more pleasing to the writer.

Research Hardware

The lines and dots that make up the display elements are generated by an Evans and Sutherland LDS-2, modified to drive beam-penetration color CRTs. Each color CRT measures 17.7×17.7 cm (7×7 in.). An SEL-840 computer interfaces with the LDS-2 to generate aircraft dynamics, navigation and guidance equations, and performance recording.

FEATURES OF THE INDIVIDUAL CCD DISPLAYS

The CCD concept as outlined so far is quite simple. However, when the amount of specific information that could go on each display is considered, along with the different possible forms that could be given to each piece of information, it is clear that the implementation of CCD could become complex. In the following description of the individual CCD displays, only one form of selected information is described. It is to be understood that changes will be made to accommodate the requirements of specific experiments, and the purpose of these experiments will be to seek better forms of the displays.

Vertical Situation Display

The Vertical Situation Display is the primary display for aircraft attitude. Since everything is referenced to the direction of flight, the center of the display can easily become overly cluttered with aircraft symbol, horizon line, pitch marks, runway symbol and other aiming points. For this reason everything that might logically go on this display cannot be accommodated at the same time. One configuration of the VSD is shown in figure 3. Element color assignments given in the text below are summarized in Table 1.

This method of showing the attitude situation is fairly standard. The combination of aircraft symbol (fixed), horizon line, and roll angle marker show a 10° left bank and zero degree pitch angle. These elements will all be shown in red.

The ground plane is differentiated from the skyplane by a perspective dot pattern. The rate of downward motion could be programmed to be a function of forward velocity and/or altitude and may be studied at a later time. It is believed, however, that the most important function of these dots is the

ground-plane/skyplane differentiation (ref. 5) and secondarily the general streamer effect (ref. 6). Altitude and velocity coding would encounter range problems; for example, the dots would be so far apart close to the ground that the visual illusion of the ground plane would be lost. The ground plane dots will be yellow.

The performance information that will have to be read most precisely during critical maneuvers surrounds the central attitude display. The altitude position reading on the right has a natural up/down relationship on this display. Also the heading readout at the top of the display has a natural right/left relationship. Airspeed has no natural position correspondence so the standard population stereotype, reading upward for larger values, was adopted (ref. 7). Heading, altitude, and airspeed are each read as a combination moving tape and digital readout, taking advantage of the best features of both. Digital readouts can be read more quickly and accurately than an analogue readout, but are poor for rate judgments. A moving tape provides rate and lead information. In operation, the moving tape numbers are blanked from the digital readout box so that the visual effect is that of the tape moving behind the box. For this simulation there will be provision for choosing either moving tape or digital readout separately before beginning a flight.

The rate of change of heading and altitude, more commonly known as turn rate and instantaneous vertical speed indication (IVSI), respectively, are displayed adjacent to the appropriate moving tape. Turn rate is normally shown in terms of a standard $3^\circ/\text{sec}$ turn (although for STOL aircraft this will probably need modification). In like manner the IVSI will be scaled for one or two standard sink and climb rates. If needed, a speed command or error bug will run along the airspeed tape. All these elements on left, top, and right of the VSD will be green.

Two pieces of information, flight path angle (FPA) and potential flight path angle (PFPA), have been combined into one symbol so that the relationship between the two pieces of information cannot be lost among other symbols on the display. In figure 3, the point between the two tips of the FPA marker is the actual direction of aircraft flight at a given moment. This point is also called the aiming point and a line extending from the aircraft toward this point is called the velocity vector. This symbol can be used to show flight path angle relative to the horizon or to any spatially located point such as a 3-D waypoint, runway threshold, or another aircraft. Flight path can be computed relative to the ground or relative to the air mass. At present there are arguments pro and con for each of these frames of reference. First emphasis will be on pilot interpretation problems with different simulated weather conditions. Green is the color from the inner tips of the symbol to the pivot or bending point.

The PFPA is referenced to the FPA. When the PFPA is level with the FPA, the acceleration along the aircraft flight path is equal to zero, therefore speed is constant. If PFPA is above FPA, the acceleration is positive and the speed will increase; if PFPA is below the FPA, acceleration is negative and the speed will decrease. These two indicators make the effect of changes in throttle setting, flaps, landing gear, etc., immediately apparent to the pilot. As an illustration of the use of these two display elements consider the

example shown in figure 3. The potential flight path is shown as being 8° below current flight path. Without a power change, a continuation of the flight path shown will lead to a reduction in airspeed. The pilot can use the information to increase throttle until the potential flight path reads the same value as for flight path, thereby maintaining current flight path and airspeed. Or, continuing with the example in figure 3, the pilot can maintain current airspeed without changing thrust by pitching down until the flight path matches the potential flight path. Because potential flight path is a directly controlled variable, the "flat" portion of the symbol is in red.

Not shown in figure 3, but planned for evaluation are waypoint guidance, runway and touchdown point, and a method for showing a 3-D perspective of desired flight path; for example, tunnel or channel display (ref. 8).

Side Vertical Situation Display

The Side Vertical Situation Display clearly and unambiguously relates present aircraft altitude to future altitude requirements (fig. 4). The aircraft symbol (red) remains fixed at the altitude digital readout box (green). This accomplishes two purposes: (1) the aircraft altitude reference is explicitly established and (2) a required second altimeter is provided. The altitude on the VSD is from radar and the altitude on the SVSD is barometric. The operation of the moving tape/digital readout is the same as described for the VSD. To enhance terrain altitude awareness, significant terrain features (yellow) can be shown referenced to the moving tape. Logic will have to be provided to change these features as a function of lateral displacement from desired ground track.

Flight path angle (green) and potential flight path angle (red) are accurately read against an expanded angle scale (3:1 in fig. 4). The vertical relationships are the same as previously explained for the VSD. The aircraft symbol rotates about its midpoint to indicate aircraft attitude.

An IVSI digital readout (green) in the upper left corner supplies absolute vertical speed information, supplementing the analogue readout on the VSD. An arrow appearing above or below the box reinforces the sign information regarding up or down velocity of the aircraft. There is a ± 50 ft/min dead band about zero ft/min so that the arrow is not continually flipping over when the aircraft is flying straight and level. In keeping with the philosophy of relating quantitative information to qualitative information this vertical speed information should be closely related to the flight path angle or the aircraft symbol. Initial attempts to do so resulted in excessive clutter and loss of other information. As with all items on these displays its final form is yet undetermined.

A segmented line (yellow), moving toward the aircraft symbol, indicates the desired vertical track. Relevant tags are shown at waypoints, marker beacons, etc. Vertical and horizontal scaling must be compatible with the flight path angle scaling.

Horizontal Situation Display

The Horizontal Situation Display (fig. 5) relates the aircraft (red) to its geographic position. This may be shown as aircraft position relative to a desired course line, navigation aids, waypoints, runways, or prominent geographic features, all of which would be shown in yellow.

The horizontal projection of the velocity vector or flight path (called a trend vector by Boeing), the range altitude symbol, and ground speed and wind-speed vectors would be green. The range altitude symbol shows the point at which the next waypoint altitude will be reached if the present vertical component of the velocity vector is maintained.

If this display is to be used for manual control the lateral track error can be expanded by some factor and shown by a bar parallel to the aircraft, as if a portion of the guide line had been cut out and expanded.

Sufficient work has been done to show the utility of a predictor on the HSD (refs. 9, 10) so that an evaluation of a predictor (not shown in fig. 5) will be part of this work. (This may eventually include evaluation of predictors on the VSD and SVSD as well.) Also not shown, but candidates for HSD presentation, are time slot information for 4-D navigation and symbols showing other aircraft for traffic situation information (refs. 11, 12).

Coordinated Cockpit Display Evaluation

The goal of the CCD concept is to present flight information explicitly in its situational context. The advantages and disadvantages of this approach remain to be studied. In the first simulator study, pilots will manually fly a complex, decelerating landing approach with go-around at 200 ft before touchdown. Using this task, pilot performance will be compared when using the CCD or standard instruments. Pilots will be interviewed for opinions, comments, suggested changes, and additions or deletions.

As display ideas evolve it is expected that various configurations of the CCD will be compared so that new ideas on display content and form can be evaluated. In a parallel effort the CCD will also be integrated into a full mission simulation and evaluated in the larger context of complex navigation with an air traffic control system.

REFERENCES

1. Wempe, Thomas F.: Flight Management-Pilot Procedures and System Interfaces for the 1980-1990's. AIAA Paper No. 74-1297, Nov. 1974.
2. Baty, Daniel L.: Rationale and Description of a Coordinated Cockpit Display (CCD) for Aircraft Flight Management. NASA TM in process.
3. Warner, John D.: Advanced Controls and Displays for Future Commercial Aircraft Operations. AIAA Paper No. 70-938, July 1970.
4. Air Force Manual 51-37, *Instrument Flying*, Dept. of the Air Force, Washington, D.C., 15 Nov. 1960.
5. Lincoln, R.; Palmer, E.; and Wempe, T.: Effects of Pictorial Display Configuration on the Frequency of Control Reversals during Aircraft Landing Approaches. NASA TM X-62,191, Oct. 1972.
6. Gibson, J. J.: *The Perception of the Visual World*. The Riverside Press, 1950.
7. McCormick, Ernest J.: *Human Factors Engineering*. McGraw-Hill, 1964.
8. Wilckens, V.: Improvements in Pilot/Aircraft-Integration by Advanced Contact Analog Displays. Proceedings of the Ninth Annual Conference on Manual Control, May 23-25, 1973, M.I.T., Cambridge, Mass., pp. 175-192.
9. Kreifeldt, J. G.; and Wempe, T.: Pilot Performance During a Simulated Standard Instrument Procedure Turn With and Without a Predictor Display. NASA TM X-62,201, Jan. 1973.
10. Baty, D. L.: Evaluating a CRT Map Predictor for Airborne Use. IEEE Trans., Syst., Man, Cybern (in press).
11. Technology Review, "What Role for a Traffic Situation Display?" Jan. 1975, p. 25.
12. Kreifeldt, J.; and Wempe, T.: Human Decision Making in Future ATC Systems: Comparative Studies in Distributed Traffic Management. Presented at the 1974 International Conference of the IEEE Systems, Man and Cybernetics Society, Dallas, Texas, Oct. 1974.

TABLE 1.- Display Element Color Assignments

Display elements	Red	Green	Yellow
VSD			
Horizon line and pitch marks	X		
Aircraft symbol	X		
Roll angle	X		
Ground plane dots			X
Altitude tape and digital readout		X	
Airspeed tape and digital readout		X	
Heading tape and digital readout		X	
Turn rate		X	
IVSI		X	
Flight path (FPA)		X	
Potential flight path (PFPA)	X		
Waypoint guidance (not shown)			X
Runway (not shown)			X
Tunnel or channel (not shown)			X
SVSD			
Aircraft symbol	X		
Altitude tape and digital readout		X	
Terrain features (not shown)			X
Flight path (FPA)		X	
Potential flight path (PFPA)	X		
Angle scale		X	
IVSI		X	
Desired vertical track			X
Waypoints, beacons, etc.			X
HSD			
Aircraft symbol	X		
Flight path		X	
Range altitude		X	
Ground/windspeed vectors		X	
Desired course line			X
Expanded error bar			X
Navigation aid			X
Waypoint			X
Runway			X
Obstructions			X

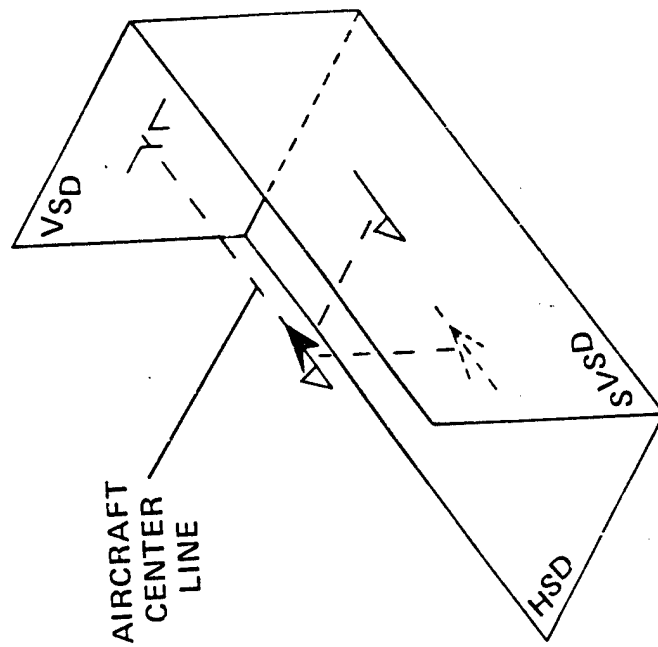


Fig. 1. Three orthogonal planes of aircraft situation.

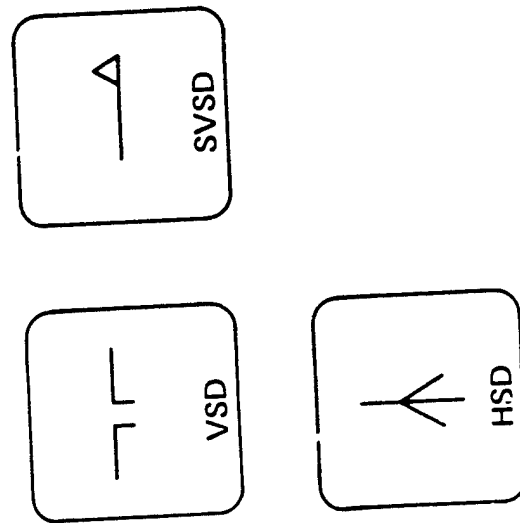


Fig. 2. Position of three displays in aircraft instrument panel.

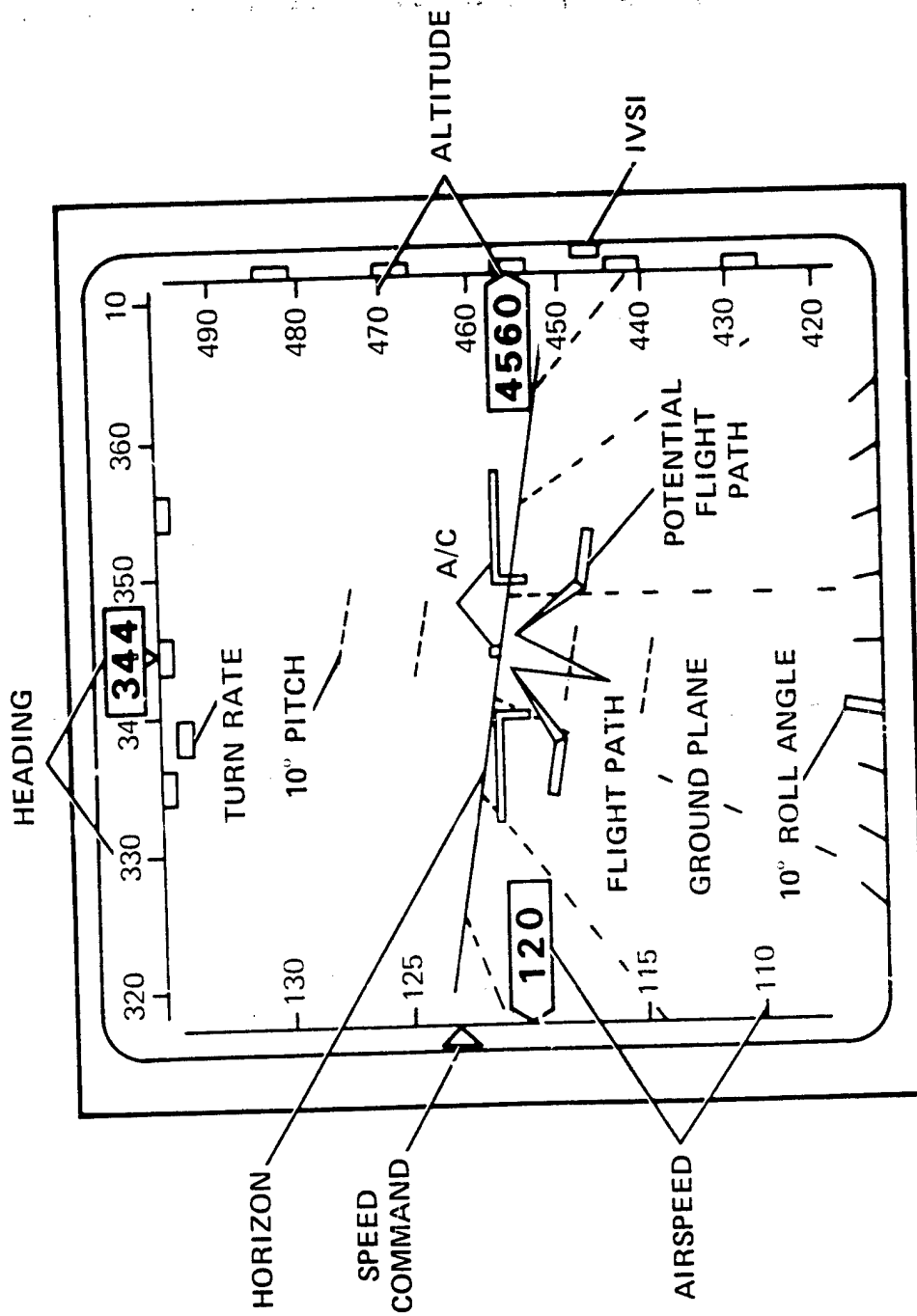


Fig. 3. Vertical situation display.

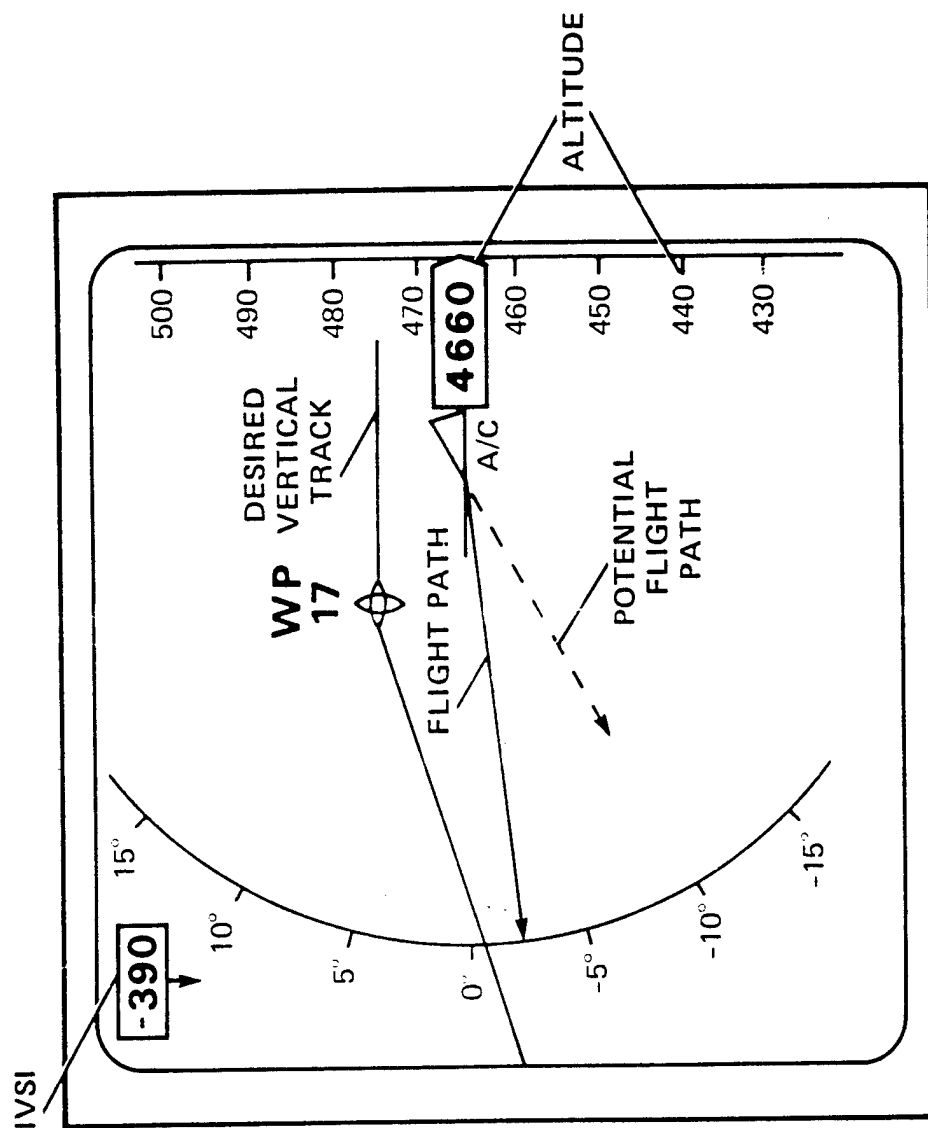


Fig. 4. Side vertical situation display.

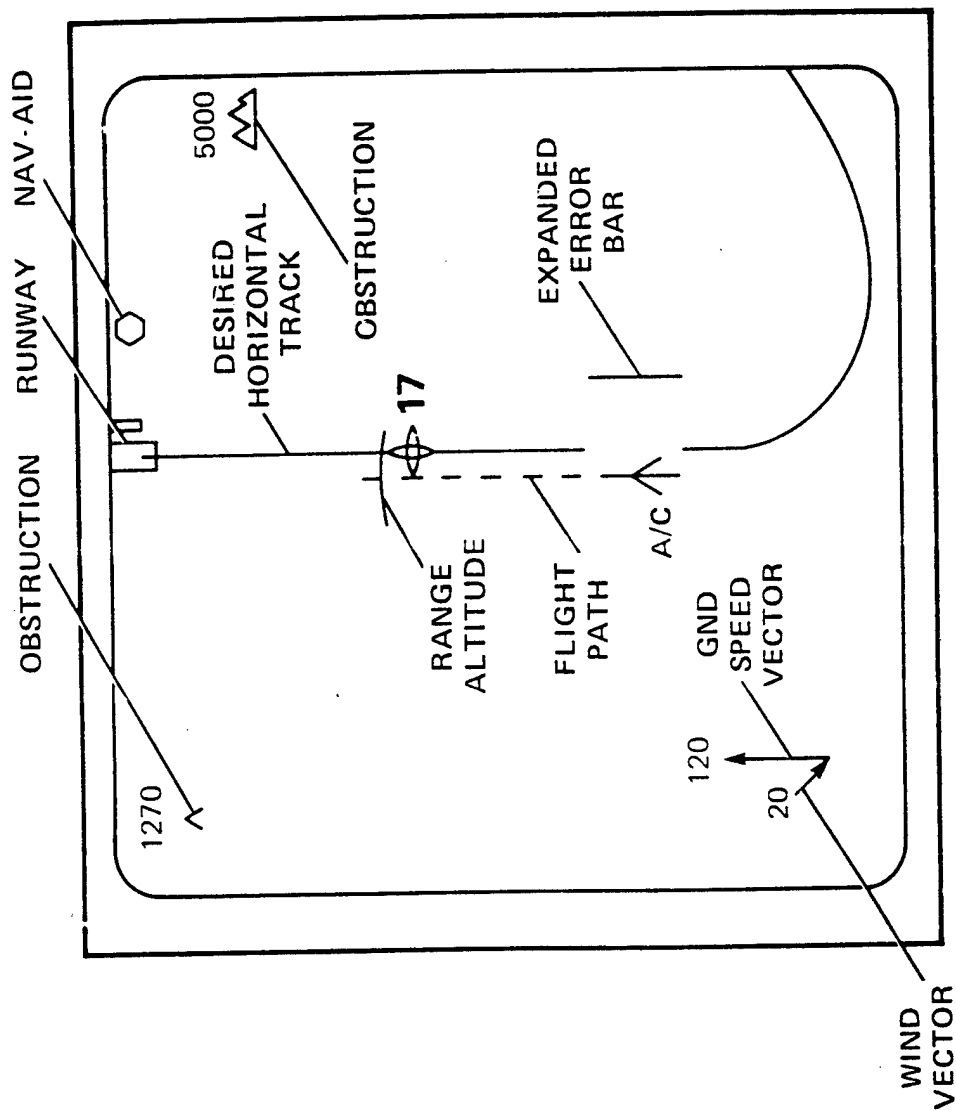


Fig. 5. Horizontal situation display.

IMPLICATIONS OF A MIXTURE OF AIRCRAFT
WITH AND WITHOUT TRAFFIC SITUATION DISPLAYS
FOR AIR TRAFFIC MANAGEMENT

J. Kreifeldt; L. Parkin; P. Rothschild
Department of Engineering Design
Tufts University, Medford, MA 02155

T. Wempe: Man-Machine Integration Branch
NASA-Ames Research Center
Moffett Field, CA 94035

Presented at the 12th Annual Conference on
Manual Control, 1976

SUMMARY

A mixture of aircraft (A/C) with and without traffic situation displays (TSD) was simulated to ascertain its effects on distributed air traffic management in the terminal area.

A particular type of distributed air traffic management (sequencing) was used as determined from previous experiments with the nonTSD aircraft being vectored. This mixed condition consisting of one A/C simulator without a TSD and two with TSD was compared to a baseline vectoring condition in which all 3 simulator A/C lacked TSD.

The three simulator A/C and four or five computer simulated A/C were embedded in a terminal area traffic problem with as much realism as possible. All A/C were considered to be STOLcraft.

Analyses were made of flight performance measures, verbal communications and subjective evaluations by the professional pilots and controllers who served as subjects.

The analyses favor the TSD equipped A/C and the distributed mode of management permitted by this cockpit capability. However there are indications that an A/C without a TSD in a TSD environment may require or receive considerably more controller attention and pilot disfavor than when it is in an all vectored environment. This may imply that TSD and nonTSD A/C should be segregated and controlled accordingly.

INTRODUCTION

A number of experiments and reports (1, 2, 3) have now documented the considerable potential that Traffic Situation Displays (TSD) in the cockpit hold for increasing the safety, orderliness and expeditiousness of terminal area air traffic management without increasing pilot or controller workloads. The controller's verbal workload in fact can be considerably reduced⁽⁴⁾ by a type of distributed management in which the TSD A/C are issued sequence orders only which they then achieve via their TSD without need for any vectoring as in VFR (Visual Flight Rules.)

However, the experiments to date have not treated the very reasonable possibility of a mixed traffic display environment in which some A/C have TSD capability and others do not. This situation could arise by failure of the TSD in an A/C. It could also arise through the more common possibility that not all A/C will have TSD due to cost or other evolutionary factors thus producing a mix of A/C on a traffic information basis. Just as a mix of A/C speed types poses special management problems and techniques it might also be supposed that a mix of A/C information types might have its own impacts on efficient management.

In order to investigate the effects of a mixed TSD environment, an experiment was performed at NASA-ARC utilizing the Air Traffic Control simulation facilities and programs developed in the Man-Machine Integration Branch. This facility permits studying the group interactive aspects and performance of terminal area traffic management based on exploitation of the 360° view of surrounding traffic made available in the cockpit by a Traffic Situation Display.

A previous study⁽¹⁾ showed that distributed management of 3 A/C equipped with TSD was not only possible but also generally superior in a number of respects to the ground centralized vectoring method of control. In particular, a form of distributed management termed sequencing appeared superior to an alternative form of distributed management. In the sequencing mode, the controller issued only sequence order to each TSD equipped A/C leaving them to achieve this order in VFR fashion thus vastly reducing controller verbal workload without any substantial increase in the pilots' workloads. Therefore for this study, the sequencing mode of distributed management was used for the TSD equipped A/C. All non TSD A/C were vectored.

Two basic conditions were simulated in the present experiment. Vectoring as a condition was characterized by all three simulators lacking traffic information and thus requiring vectors. Nonvectoring was identified with the mixed information condition in which two of the simulators had TSD but the remaining one did not. In a sense, the mix could be viewed as an independent variable along a continuum with the

ratio possibly determining the best form of traffic management. The mixed ("nonvectoring") condition was thus a simultaneous VFR-IFR environment.

The two basic conditions (vectoring, nonvectoring) were compared and within the nonvectoring (mixed) condition, the two specific management techniques (Sequencing, Vectoring) were further compared. Comparisons were based on flight performance measures, verbal communications and subjective evaluations obtained after each run and after the experiment. Three groups of three pilots and two controllers per group were used to test for group specific differences in the results. All subjects were practicing professionals.

METHOD

1. Task

The basic task is referenced to the map shown in Figure 1.

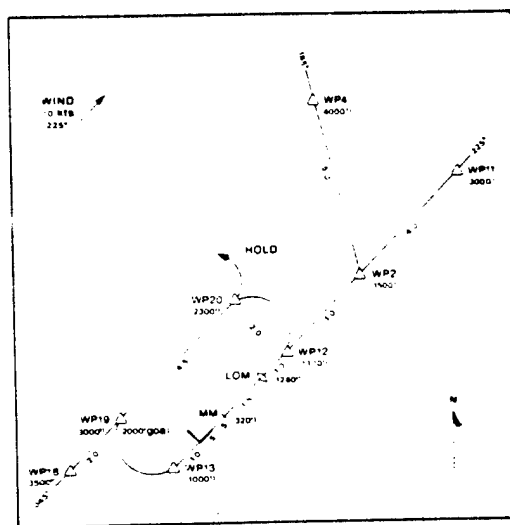


Figure 1. Ground Projection of the Task Layout. Distances are given in Nautical Miles. A STOL Terminal Area is simulated.

The Locations of the Outer Marker (LOM) and Middle Marker (MM) were appropriate for a STOL airport. There were three approaches to the airfield and a go-around (goa) route to be used in case of a missed approach.

Suggested altitudes are given at each waypoint (WP) for traffic purposes although once past WP 12, the altitudes shown are appropriate for a 6° STOL approach which all craft were required to execute.

Computer generated A/C were introduced at WP 11 every two minutes on the average with a 5 second standard deviation. Once introduced, each A/C followed a standard altitude and speed profile appropriate for its type. The controllers could modify only the speed of these craft through keyboard input. Controllers could "hold" the computer A/C at WP 11 and subsequently resume their introduction through appropriate keyboard commands.

The piloted simulator A/C were initially introduced anywhere between or on the 045° and 165° approaches to the terminal area with appropriate headings and speeds. Pilots had aileron, elevator and throttle control of their craft.

The basic task required that the 3 simulator craft be inserted between the computer A/C (scheduled to cross WP 12 every two minutes with some variation.) A basic rule limited spacing to a minimum of 1nm or 60 second between A/C crossing the middle marker. (A fuller description of the problem and subject instructions is given in reference 5.)

Two management conditions for accomplishing this task were compared. In the first condition (vectoring) all 3 A/C were without TSD. Each pilot saw the map of Figure 1 and his own A/C position on it. Because no visual traffic information was available to the pilots they had to be vectored. In the second condition (nonvectoring), two of the three simulators had full 360° visual traffic information while the third remained as in the first condition with the map and only his own position shown on it. Thus two of these A/C (with TSD) could operate in the sequencing mode while the third A/C (without TSD) required vectoring. A common voice circuit was used between all pilots and controllers. Each pilot had a vertical situation display identical to that used in previous studies and a description may be found in Reference 2. A 30 second path predictor was always present on own A/C only.

The basic objective of this study was to determine the impact such a heterogeneous mixture of A/C with and without traffic information would have on terminal area traffic management when compared with a baseline vectoring condition.

2. Measures

Three types of measures were obtained in this experiment. Objective (flight performance) measures of variables such as aileron, elevator and throttle activity in the simulators; final airspeed, heading, glideslope, etc., errors, pitch and roll rates and others: verbal measures based on

tape recordings of the pilots and controller exchanges: and subjective measures obtained from questionnaires completed after each run and after the total experiment.

The experiment was replicated on three separate groups of three pilots and two controllers per group. Three runs of each of the two experimental conditions were obtained from each group. A run lasted 20 to 25 minutes and each group received approximately 4 hours of practice under both conditions. Practice and test runs were made on separate days. All 15 subjects were current airline pilots or air traffic controllers from the San Francisco-Oakland region.

To reiterate, in the Vectoring condition, no A/C had a TSD and thus all A/C required vectoring. In the Nonvectoring condition, the two A/C with traffic displays received sequence order information only from the controller and were left to achieve that order via their TSD and any verbal communication necessary between the pilots. The single A/C without a TSD had to be vectored by the controller among the A/C flying "VFR." Thus the nonvectoring condition had both VFR (TSD equipped A/C) and IFR (nonTSD equipped A/C) traffic.

RESULTS

The results are highlighted here with more detailed analyses to be found in reference 5. Results are presented according to the three types of measures.

1. Objective Measures

A. Comparisons

A battery of seventeen objective performance measures was recorded on each of the 3 piloted simulated aircraft. These variables are not statistically independent of each other. However, an assessment of the number of independent components present in the battery will depend, to some extent, on the results of the present analyses. Application of principal components and factor analytic-procedures are therefore postponed to a later report.⁽⁵⁾ Sixteen of the variables may be thought of as belonging to four general groups as follows:

Group 1 - Landing Accuracy
FAE: final airspeed error
FGE: final glideslope error
FHE: final heading error
FLE: final lateral error

Group 2 - Manual Workload
AA: aileron activity
EA: elevator activity
TM: throttle movements
TM/DT: throttle movements
per unit of flight time

Group 3 - Expedience/Economy
HPL: horizontal path length
DT : duration of flight
STM: speed total maneuvering
ICT: intercrossing time

Group 4 - Passenger Comfort
RRA: roll rms acceleration
PRA: pitch rms acceleration
TAC: throttle average change
STM/DT: speed total man. per
unit of flight time

All measures in the Landing Accuracy group were made as the aircraft passed the middle marker (MM on Figure 1) during the final approach.

In the second group, the aileron and elevator activities are rms values of these controls during the simulated flight. Throttle movements (TM) which resulted in speed changes of five knots or more were counted for this variable. The variable TM/DT is the throttle movement count divided by the duration of the flight.

The third group of output variables includes the total length of the aircraft ground track (HPL) and the time duration of the simulated flight (DT). These two variables are highly correlated. Speed total maneuvering (STM) is the sum of all speed changes during the flight. Intercrossing time (ICT) is the time interval between a given aircraft following the one ahead of it across the middle marker on the final approach. (Subjects were instructed to maintain an intercrossing time as close to sixty seconds as possible.) The lowest mean ICT for any of the ATC management schemes was in excess of the sixty second guideline.

The final group of output variables describes the aircraft motion as it may affect passenger comfort. The first two variables are the rms values of roll and pitch accelerations. The average throttle change (TAC) was computed using only throttle changes which would result in speeds in excess of five knots and, therefore, reflects the more gross and perceptible speed changes. The speed total maneuvering was divided by the flight duration to give the last variable in this group, STM/DT.

For each of these sixteen variables, the lowest value is also the most preferred value. A seventeenth output variable is the average speed of the aircraft (AVSPD). There is no obvious optimal value for average speed and since there was no significant difference in this variable among the several ATC management schemes, it was not included in the summary of results to follow.

An a priori summary of the objective performance results is presented in Figure 2, showing four ATC management categories. The first category, vectoring, includes 44 completed flights. The cockpits of simulators were equipped with displays showing the position of that aircraft only in relation to the map. The next three categories involve nonvectoring condition flights. The first of these included 35 completed flights in which the simulators were equipped with traffic situation displays showing the positions of all airborne aircraft

to the pilots. The second of the nonvectoring categories included the 16 vectored flights of simulators with no TSD's. Experimental runs in the nonvectoring condition were heterogeneous since simulators with and without TSD's were in the air at the same time in a ratio of about 2:1. The last category of the nonvectored condition flights included all 51 flights of the two previous categories (both with and without TSD's.) This category reflects the overall performance of the two to one mix of TSD and non TSD aircraft for comparison with the totally vectored condition flights.

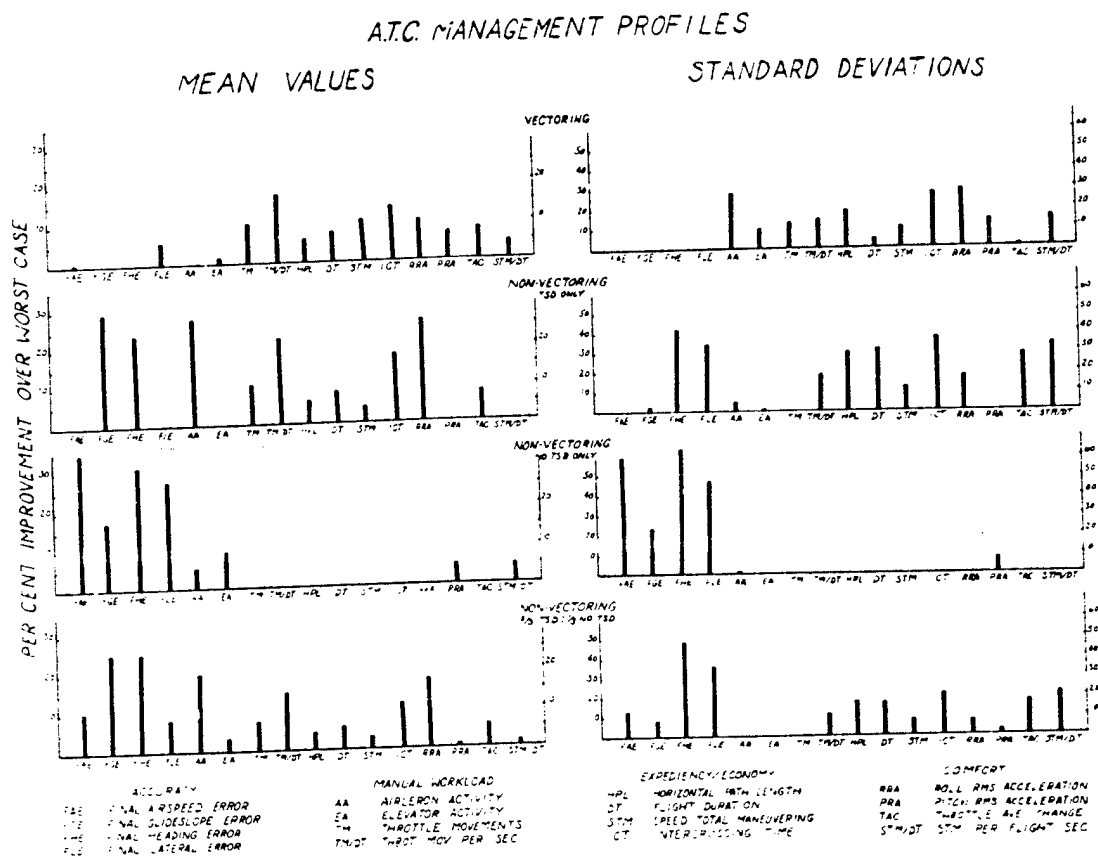


Figure 2 A Relative Comparison of the Means and Standard Deviations of the 16 Objective Measures for Four Types of Flights. Percentages shown are Relative to the lowest Value which is not shown.

When a flight involved a missed approach or "go around", the data describing the portion of the flight after the aircraft passed the middle marker for the first time is not included in the results and analysis. This data was omitted since the high throttle, aileron, and elevator activities and consequent aircraft accelerations associated with recovering from the missed approach tended to produce markedly bimodal distributions for these variables.

The profiles of Figure 2 compare the mean values and standard deviations of the first sixteen output variables. The ATC management category which produced the highest mean or standard deviation for a given variable is taken as the worst case for that variable. The vertical bars indicate the per cent improvement achieved by the given management schemes over the worst case for that variable. For instance, nonvectored TSD's produced the worst results for the mean value of final airspeed error. The vectored flights produced results which were .5 per cent better (lower). Nonvectored condition flights without TSD's (i.e. vectored A/C in the mixed condition) achieved a 35 per cent improvement over the TSD flights on this variable and the combined result for all nonvectored condition flights was about 10 per cent better than the TSD category alone.

Figure 2 shows that the vectoring condition produces poor results in the final error measures both with respect to mean values and standard deviations. Vectoring also resulted in consistent but poor performance with respect to aileron activity. However, the remainder of the manual workload measures, as well as the expediency/economy measures and passenger comfort measures constituted improvements over the respective worst cases. The nonvectored TSD flights produced poor results in final airspeed and final lateral errors, although glideslope and heading errors were low. This category produced the best results with respect to aileron activity and consequent roll acceleration while requiring the greatest amount of elevator activity and pitch acceleration. All other performance measures seem to follow the vectoring profile. The nonvectoring condition runs without TSD's (i.e. vectored flights in the mixed condition) outperformed all other categories on the final error measures. However, this category produced the worst results for almost all other measures.

Figure 3 shows a set of pair-wise comparisons of the various ATC management categories for aiding in comparing flights on the 16 measures. The first comparison contrasts vectoring with the combined nonvectoring category (i.e., no distinction for TSD, nonTSD.) Nonvectoring appears to outperform vectoring in the final error measures while the reverse is true for all but two of the other measures. The aileron activity favored nonvectoring and was the only measure significant at the .01 level. Comparison of the standard deviations on these two categories showed nonvectoring to produce more consistent results in the final error measures. F tests showed the heading and lateral errors to be significantly more consistent (at the 0.1 level) for nonvectoring. The variance in aileron activity and roll acceleration was significantly lower (at the .05 level) for the vectoring flights however.

PAIR WISE COMPARISON OF ATC MANAGEMENT CATEGORIES MEAN VALUES STANDARD DEVIATIONS

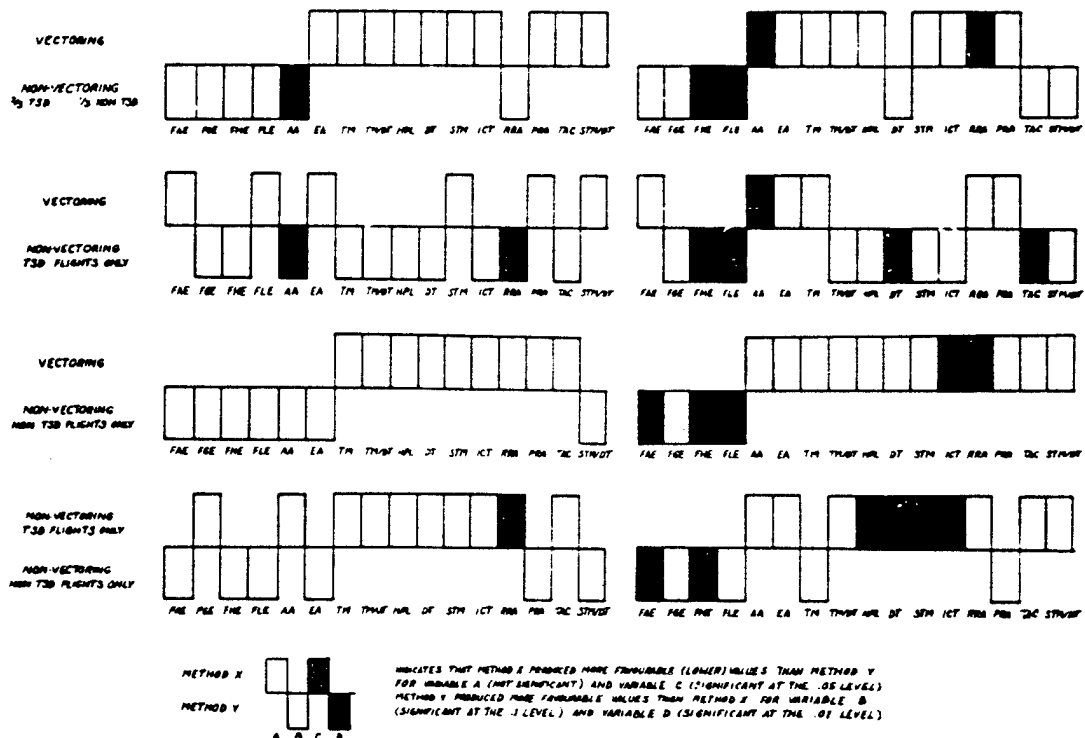


Figure 3 Means and Standard Deviations of the 16 Objective Measures Compared for the Four Types of Flights on a Pair-Wise Basis. The Favoring Direction of a Measure is Indicated. Levels of Statistical Significance are Shown.

Comparison of the vectoring and nonvectoring (TSD) flights generally favored the latter category. The nonvectored flights had significantly lower aileron activity and consequent roll acceleration. The variance of FHE, FLE, DT, TAC, and STM/DT was significantly less for the nonvectored TSD flights while the variance for aileron activity significantly favored the vectored aircraft.

Comparison of vectoring condition flights with the nonvectored condition non TSD (i.e. vectored) flights showed the latter to perform better with respect to final error measures and aileron and elevator activity, while vectoring was favored on most other measures. No comparison of mean values between these categories was significant. The nonTSD vectored flights in the nonvectoring condition produced less variance in all of the final error measures and three out of four of these comparisons were significant beyond the .01 level. Vectoring condition flights produced lower variances in all other measures although only aileron activity, roll acceleration and intercrossing time were significantly lower.

Finally, the nonvectored TSD flights are compared with the nonTSD (vectored) flights in the nonvectoring condition. The only significant differences in mean values show TSD flights to be superior with respect to aileron activity and roll acceleration. The nonTSD flights have lower variances in all of the final error measures (two out of four were significant) while the TSD flights had significantly lower variances in all of the expedience/economy measures. The variance in throttle average change also favored the TSD flights.

A brief summary comparing the conditions on the relative favorableness of each measure (mean and standard deviation) shows the following:

- O Vectoring condition superior to nonvectoring condition overall.

Vectoring showed more favorable biases than nonvectoring although the most significant differences favored nonvectoring as did the final error measures. Intersecting times favored vectoring.

- O TSD nonvectoring superior to vectoring.

The truly nonvectored flights (those with TSD) compared to the flights in the vectoring condition showed more favorable biases (10-to-6 for both means and standard deviations) including intersecting times. In addition, the 7 significant differences all favored the TSD flights.

- O Vectoring condition superior to vectored nonTSD.

The flights in the vectoring condition showed more favorable biases than the vectored flights (noTSD) in the nonvectoring condition, most noticeably in the standard deviations (12-to-4) although again the most significant differences favored the nonvectoring nonTSD flights as did the final errors. Intersecting times favored the vectoring condition although not significantly.

- O TSD nonvectoring superior to vectored nonTSD.

Within the nonvectoring condition, the TSD (nonvectored) flights showed more favorable biases than the nonTSD (vectored) flights including intersecting time. However the final errors were smaller and more consistent for the nonTSD.

Thus, based solely on objective measures, the TSD (nonvectored) flights were superior to vectored flights whether or not all the flights in the condition were vectored or just those flights lacking a TSD. On the other hand those flights that were vectored in the mixed (nonvectoring) condition were somewhat better on final errors.

B. Multidimensional Analyses

A series of multivariate analyses of variance (MANOVA) and discriminant analyses were performed on the data using the ATC management schemes as the independent or grouping variable. Thus, there were three

groups: vectored condition flights, nonvectored condition flights with TSD's, and nonvectored condition flights without TSD's. The purpose of these analyses was to determine if there were significant differences among the 3 groups in terms of the entire set, or a smaller subset, of the output measures. If such differences are present, as a priori analysis implies, discriminant variates should express these differences in a readily interpretable fashion.

The first step in the analysis was to break down the total variance in the measurement battery of seventeen variables into a partition of variance which contained differences among the three groups and a partition which included variation within the groups. The hypothesis H_1 , equal within dispersion (variance-covariance) matrixes for the three groups, was then tested using an F-distributed transformation of Box's M statistic⁶. This hypothesis was retained thus validating application of the MANOVA procedure. An F-distributed transformation of Wilk's Lambda statistic⁶ was then used to test the hypothesis, H_2 , that the column vectors of mean values on the members of the measurement battery were the same for all three groups. It was found that this hypothesis could be rejected with a probability of false rejection less than 0.1.

Inspection of univariate F ratios for the individual measurement variables enabled evaluation of the discriminating power of the various variables. Two of the less useful variables were eliminated to produce a smaller subset of 14 variables and both hypotheses were retested for the smaller measurement battery. Hypothesis H_1 was again retained and hypothesis H_2 was rejected with a smaller probability of false rejection. The number of variables was selectively and finally reduced in ensuing analyses to 3.

The eight member battery included the variables HPL, STM, AA, RRA, FGE, FAE, DT, and ICT and enabled rejection of H_2 with a probability of false rejection at less than 0.01. The discriminant analysis performed on this battery produced two statistically orthogonal components of the battery, both of which provide significant discrimination among the groups. The difference among the three group means on the first component was significant beyond the 0.005 level. This component is defined in terms of the standardized scores of the eight variables and is given in standardized form by the following equation:

$$D_1 = .067 \text{ HPL} - .328 \text{ STM} + .592 \text{ AA} + .372 \text{ RRA} \\ - .323 \text{ FAE} + .377 \text{ FGE} + .320 \text{ DT} + .397 \text{ ICT} \quad (1)$$

The level of significance of differences among group means on the second component was between 0.05 and 0.02. This component is given by the following equation:

$$D_2 = -.025 \text{ HPL} + .491 \text{ STM} - 1.299 \text{ AA} + 1.132 \text{ RRA} \\ - .221 \text{ FAE} + .109 \text{ FGE} + .307 \text{ DT} + .481 \text{ ICT} \quad (2)$$

The structure matrix for these components was computed using the total (before partitioning) variance in the test battery. This structure matrix, as well as the communalities achieved by the structure and the univariate F ratios computed in the MANOVA are presented below.

Variable	Structure		Communalities	Univariate F Ratios
	1	2		
HPL	.173	.246	.091	.72
STM	-.056	.312	.101	.70
AA	.709	-.324	.608	6.56
RRA	.602	.129	.379	4.06
FAE	-.147	-.270	.095	.72
FGE	.231	-.226	.04	.89
DT	.217	.298	.36	1.10
ICT	.307	.334	.206	1.176

The elements of the structure matrix are product moment correlations between the variable (row) and the discriminant component (column). The square of an element is, therefore, the proportion of total variance in the variable that is used by the component. The communalities for the variables are the sums of squares across the rows of the structure and are, therefore, the proportions of total variance in the variable that are utilized in the discriminant space defined by the two components. The univariate F ratios and communalities both indicate similar orders of importance for the variables in this measurement battery. The aileron activity is the most important differentiator of the group. The other variables are RRA, ICT, DT, FGE, STM, FAE and HPL in descending order of importance.

The mean group scores on the two discriminant components are plotted in Figure 4. Each group mean is at the center of an eclipse and the ellipses define regions of the discriminant space containing about 67% of their respective populations. The vectors emanating from the origin of the space are projections of the measurement vectors on the discriminant plane indicating their directions of increase in the discriminant space, while their lengths indicate their relative importances as discriminators of the groups.

DISCRIMINANT PLANE

EIGHT VARIABLE CASE

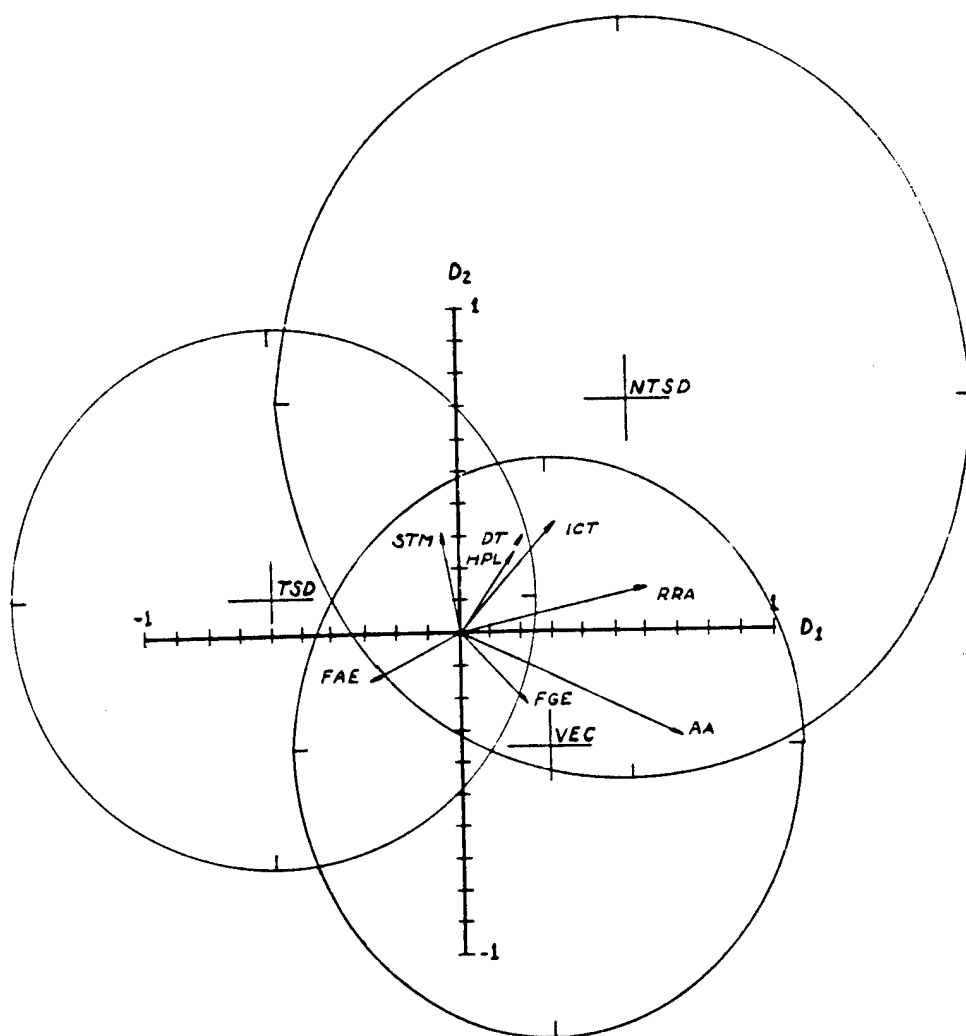


Figure 4 Discriminant Analyses of the Three Different Types of Flight. NTSD Indicates the IFR Flights in the Mixed IFR-VFR Environment. The Figure Shows a Clear Significant Difference Between these Flights. Eight Discriminatory Variables are Shown.

The MANOVA and discriminant analyses have established that the level of significance for differences in the smaller subset of eight variables is beyond the 0.01 level. At this point it would be appropriate and informative to perform a principal components analysis and factor analysis on the data to determine the number of independent components of variance present in the seventeen output variables and to find a rotation of the components which is maximally interpretable. Since the MANOVA showed significant differences among the groups, these further analyses must be performed on a correlation matrix obtained from the pooled within groups dispersion matrix. Use of this within groups partition of variance will eliminate the differences between groups and give a more accurate view of the inter-relationships among the individual variables. This analyses is being performed.

C. Conclusions from Objective Analyses

Therefore based on the above analyses, it appears that the flights that merged and made approaches from sequence order information only (i.e. via TSD) achieved distinctly different and better individual and system performances than vectored flights. On the other hand and perhaps somewhat counterintuitively, those flights that were vectored in a mix of vectored and nonvectored ones were also distinctly different from flights in which all were vectored. This perhaps indicates that controllers found it more difficult to vector A/C in a nonhomogeneous mix than when all were vectored even though the A/C "workload" judged by vectored A/C numerosity was 1/3 lower in the mixed condition than in the all vector condition.

2. Verbal Measures

The verbal communications during each run were tape recorded and later transcribed to hard copy for further analysis of the verbal workload for both vectoring and nonvectoring flight simulations. As before, the traffic situation display flights and those flights without traffic situation displays are compared, and more specifically the flights with TSD are compared to flights with the vectoring condition and to vectored flights within the nonvectoring condition. (Vectored pilots flying in the nonvectoring condition will be referred to as pilots without a TSD.)

A. Word Rates and Word Counts

Running cumulative word rates (averaged over successive 60 second intervals) are shown for pilots, controllers and the total group in Figure 5a, 5b, 5c.

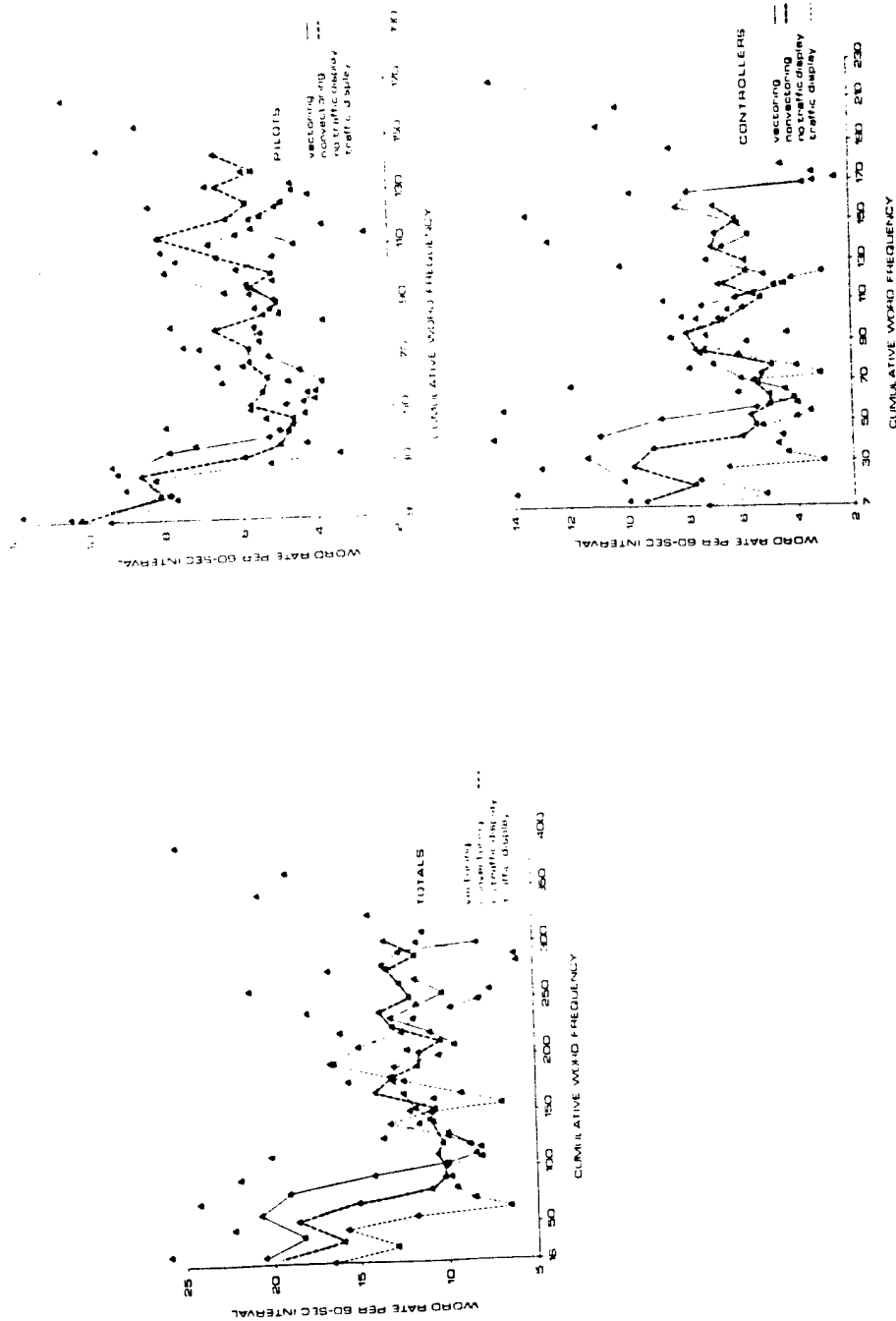


Figure 5 The Word Rates and Cumulative Word Counts for Pilots, Controllers and Both at 60 Second Intervals. The Low Verbal Workload with TSD Flights is Apparent.

In these figures, rates are plotted against the running cumulative word count permitting both to be compared. Comparisons must be made at equal numbers of plot points (60 sec. intervals). One immediate feature of the graphs shows a high initial word rate which decreases sharply followed by a subsequent increase from which point the rate continues to decrease for several minutes. While the exact behavior is no doubt task specific this same general behavior was observed in a previous experiment. (7)

Figure 5a which combines the total pilot and controller communication presents a fair comparison of the conditions. Vectoring and nonvectoring conditions had the same total word count (300) although the initial rates were lower for nonvectoring. However, where the two types of flights (TSD, NTSD) within nonvectoring are compared a large difference is apparent. Both the initial word rate and final word count were about 75% greater for the NTSD flights than for the TSD ones. This type of finding is completely consistent with previous experiments. An interesting and somewhat unexpected finding is that the word rate for the vectored flights in the nonvectoring condition (i.e. nonTSD) was considerably higher even than for flights in the vectoring condition. This phenomenon was observed for each of the three groups individually so that it appears to be fairly robust.

When just the pilot communication is looked at in Figure 5b, the differences between these 4 conditions are not as marked although again the NTSD flights had a higher initial word rate and final word count than either of the other 3 conditions. TSD (nonvectored) flights had the lowest initial word rates.

Figure 5c shows that the effects of the four conditions were most pronounced for the controllers' communication with the same general findings holding as before. The nonvectored flight (TSD) had lowest verbal workload in terms of initial word rate and total word count while the vectored flight in the nonvectoring condition (NTSD) had the highest word rate and total word count.

The group consistency effect is shown visually in figure 6 as a means of suggesting the robustness of the previous findings. Figure 6 plots the normalized differences in total word counts for each comparison. The controller groups were quite consistent across the comparisons with somewhat less consistent results for the 5 pilot groups.

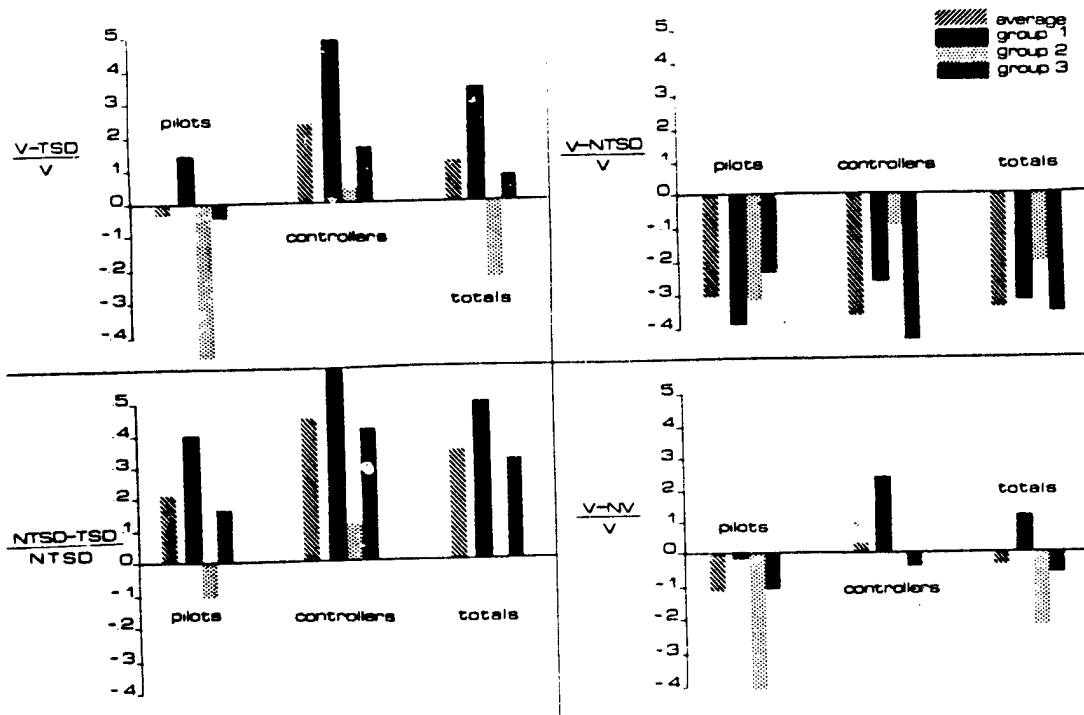


Figure 6 Pair Wise Comparison of Total Word Counts in Each Flight Type for the Three Groups.

B. Content Analyses

Content analyses are being performed on the verbal data. The results to data show that the pilots with TSD's spent significantly less time discussing direction and heading than under either of the two vectoring situations. And not too surprisingly, controllers communicating with the nonvectoring flights (TSD) made fewer references to speeds than when communicating with the vectored flights.

C. Conclusions Based on Verbal Analyses

The verbal workload for controllers is considerably reduced in both word rate and total words by taking advantage of the pilots' TSD capability. Pilots with TSD have essentially the same work rate and total word count as when they are in an "all vector" environment. Thus the lower controller verbal workload in the sequencing condition is not reflected in a higher verbal workload for the pilot. This finding is essentially the same as in a previous experiment.⁽¹⁾

The effect of a mixed traffic information environment however is surprising. The verbal workloads for both controllers and pilots being

vectoring in the mixed control environment was markedly higher than when the environment was not so mixed. This may be caused by an increased difficulty in vectoring A/C around in a space also being used in a "VFR" mode by other pilots. It might also simply reflect the increased time the controller has when some of the A/C do not need vectoring causing him to increase his communication with the remaining A/C. Analyses to determine the pertinence of this increased communication is being pursued.

In practice, this finding might suggest a necessity to segregate TSD and NonTSD equipped A/C handling each group separately on a "VFR", "IFR" basis.

3. Subjective Measures

A. Questionnaire Results

Subjects filled out evaluation sheets after each run and a final questionnaire at the conclusion of all of their runs. Figure 7 summarizes some of the averaged results obtained. Subjects were asked to indicate their assessments by placing tic-marks on an ungraduated line. Thus placement of a mark indicates not only relative ordering but strength of the assessment as well.

The overall preference for the two main conditions is shown by System Evaluation on the final questionnaire. As in a previous experiment,⁽¹⁾ controllers preferred vectoring to nonvectoring while pilots preferred the nonvectoring (i.e. mixed) environment even though they had experienced both TSD flights and nonTSD flights (in which they were vectored). The strengths of the two choices were essentially identical for pilots and controllers.

The pilots also felt that the nonvectoring condition was safer than vectoring although again controllers had the reverse opinion. However, a more detailed analysis obtained after each run shows that controllers felt the nonvectoring condition to be slightly safer than vectoring. The pilots judged the TSD (nonvectored) flights as most safe and the vectored flights in the nonvectoring environment (NT) as least safe, even compared to being vectored in an all vector environment.

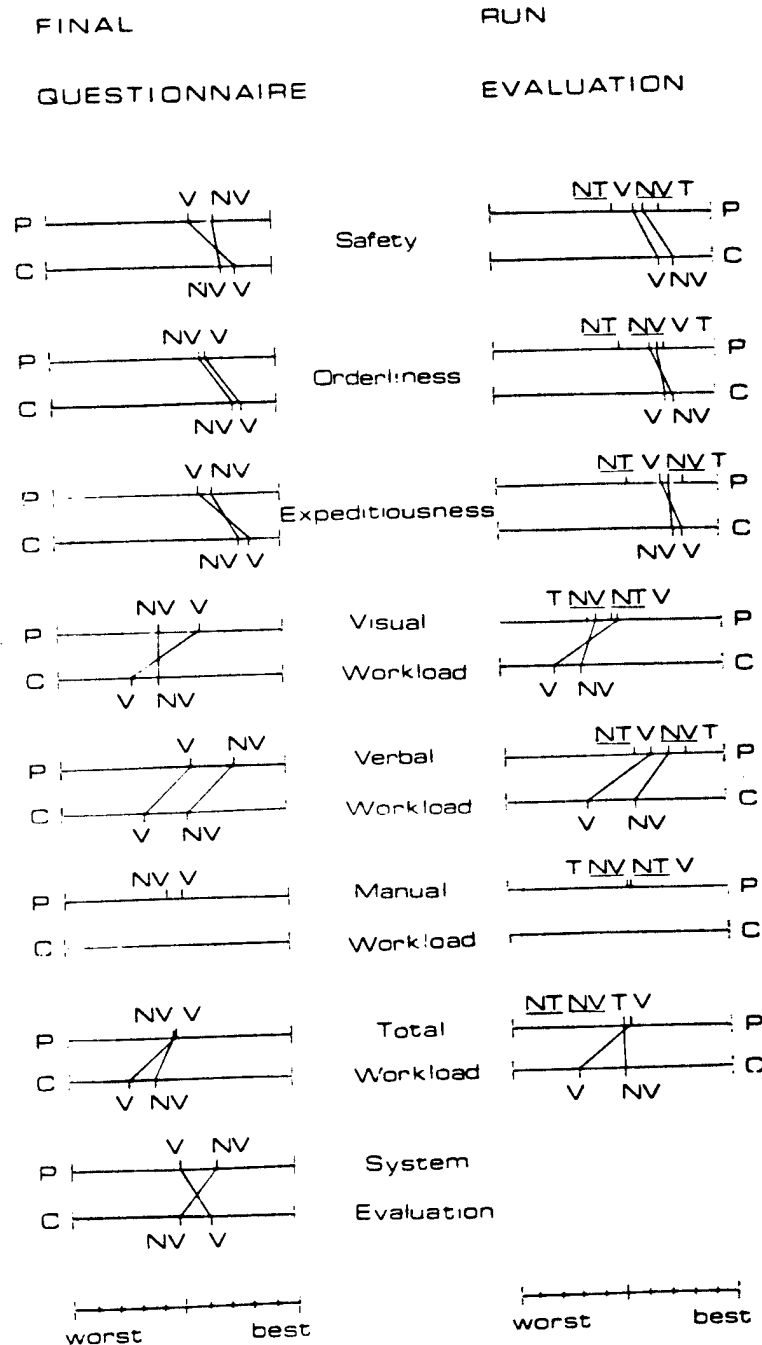


Figure 7 Pilot and Controller Averaged Responses to Questions After each Run and After the Experiment. The Final Questionnaire Compares the Two Basic Conditions. The Run Evaluation Separates TSD (T) from NonTSD (NT) Flights in the NV Condition.

In fact, pilots rated the vectored flights in a mixed environment (NT) lowest on the three primary FAA Criteria of Safety, Orderliness and expeditiousness. The fact that NTSD received lowest marks on these subjective evaluations, on the verbal measures (subjective and objective) and on many of the objective measures is a very strong finding.

Pilots felt the total workload to be essentially the same regardless of the condition (V, NV) or the type of flight made within the NV condition. Controllers quite clearly ranked the vectoring condition as having a higher (worst) verbal and total workload whether assessed after each run or in retrospect at the end of the experiment. Pilots and controllers also clearly assessed the vectoring condition as requiring a higher verbal workload.

B. Conclusions from Subjective Measures

A fairly consistent theme is that even though pilots preferred the nonvectoring condition to vectoring on the whole, they least liked being vectored in an environment where others were flying VFR. Pilots quite consistently preferred flying VFR (even in the mixed condition) to being vectored. Controllers, acknowledged that the nonvectoring condition as a whole had a lower workload and did not necessarily express any large difference between the two conditions on the three primary FAA criteria. However, in an encompassing evaluation, controllers preferred to operate in the wholly vectoring environment.

The above remarks are quite in keeping with a previous experiment particularly as to pilot-controller system preferences. Pilots want to fly VFR after a sequence order is established and controllers want a strong ground centralized system regardless of the workload decrease possible in a distributed management mode of control.

CONCLUSIONS

The conclusions offered are based on the objective, verbal and subjective analyses presented above. In considering the reality of an environment in which some A/C have TSD while others don't, it should be remembered that this situation was predicated on two possibilities, the first could be the failure of an onboard TSD in a homogeneous TSD environment while the second follows the simple realization that such a mixed environment could be quite natural based on the cost of TSD equipment.

The analyses support the primary conclusion that the mixed environment simulated can be managed effectively on a distributed management basis in which those A/C with TSD are given initial sequence orderings and then fly "VFR" while the I/C without TSD are vectored and speed controlled.

Since this mixed condition as a whole produced results at least as good as the all vectoring one (and in many cases better) it does not appear that a TSD failure on the insertion of nonTSD equipped A/C into TSD traffic has any real disruptive effect. (Of course neither one of these possibilities is necessarily advocated.)

The TSD equipped A/C operating under distributed management (controllers issue sequence, pilots fly "VFR") achieved generally better results than when they were vectored. For example, verbal communication workload was considerably lower, intercrossing spacing was closer to the desired value with less variability, etc., and this mode of control had clear pilot acceptance.

However, the nonTSD equipped A/C in the mixed environment required more verbal communication than in an all-vector environment, generally achieved lowest performance scores and was least well accepted by pilots even though from the pilots' point of view he was vectored as in the all-vector condition. This may indicate a lack of assurance by the pilot on being IFR in a VFR-IFR environment without knowledge of where other traffic is. From the controller's viewpoint, the increased verbal communication to the nonTSD A/C in the mixed environment may reflect some increased difficulty in vectoring an A/C among others flying VFR.

In the context of the present experiment, spatially segregating the TSD from the nonTSD A/C and using distributed management for TSD and vectoring for nonTSD should form the basis for a considerably better overall system than either a mixed or all-vector one.

The generally better results with distributed management for TSD in comparison to vectoring is fully in keeping with previous experiments at Ames Research Center.

ACKNOWLEDGEMENT

Funds for the support of this work have been allocated by the NASA-Ames Research Center, Moffett Field, California. under Interchange No. NCA2-OR785-501.

This support is very gratefully acknowledged as is the outstanding programming support provided by Mr. Ronald Schmeckley of PMI.

REFERENCES

1. Kreifeldt, J.G., Wempe, T.E., "Human Decision Making in Future ATC Systems - Comparative Studies in Distributed Traffic Management." Proc. IEEE International Conference on Systems, Man and Cybernetics, Dallas, Texas, October 2-4, 1974.
2. Kreifeldt, J.G., Wempe, T.E., "Future Terminal Air Traffic Management Concepts." Proc. 10th Annual Conference on Manual Control AEIT. AFFDL. April 9-11, 1974.
3. Connelly, M., "The Role of the Airborne Traffic Situation Display in Future ATC Systems." IEEE Transactions on Communications, May 1973.
4. Kreifeldt, J.G., Pardo, B., Wempe, T.E., Huff, E., "Verbal Workload in Distributed Air Traffic Management." Proc. 11th Annual Conference on Manual Control. NASA TMX-62,464, May 1975.
5. Kreifeldt, J.G., Parkin, L., Rothschild, P., The Effects of Heterogeneous Traffic Information Displays on Distributed Air Traffic Management - Final Report. NCA2-OR785-501. In preparation.
6. Cooley, W., Lohnes, P., Multivariate Data Analysis. John Wiley and Sons, 1971.
7. Pardo, B.G., Effects on Verbal Communication and Task Variables of Different Air Traffic Control Configurations. Unpublished Master's Thesis, San Jose State University, San Jose, CA 1974.

INVESTIGATING THE USE OF A MOVING MAP DISPLAY AND A
HORIZONTAL SITUATION INDICATOR IN SIMULATED
POWERED-LIFT SHORT-HAUL OPERATIONS*

By Warren F. Clement

Systems Technology, Inc., Hawthorne, California

SUMMARY

The purpose of this research-in-progress is to investigate the use which pilots make of a moving map display from enroute through the terminal area and including the approach and go-around flight phases. Various features of each of the primary STOLAND[†] displays, the electronic moving map Multifunction Display (MFD), Horizontal Situation Indicator (HSI), and Electronic Attitude Director Indicator (EADI), are used in the three phases of flight mentioned above when the STOLAND system is operated in each of three ways: a) flown in the fully automatic mode with the pilot(s) in a monitoring role; b) flown manually using flight director guidance to reduce workload and task requirements in an acceptable level; or c) flown manually using raw instrument situation data. Eye-point-of-regard and workload measurements, coupled with task performance measurements, pilot opinion ratings, and pilot comments are presented. The experimental program was designed to determine the pilots' use of the MFD in conjunction with the other displays. The measurements, ratings, and comments provide an indication of the utility of the MFD as a supplement to the HSI for improving flight safety in following curved courses and holding patterns.

INTRODUCTION

This research is designed to provide for a systematic comparison of an electronic Multifunction Display (MFD) and an electromechanical Horizontal Situation Indicator (HSI) in conjunction with other instruments (EADI, altimeter, airspeed indicator, etc.) in the NASA/Ames Research Center STOLAND digital avionics system for guidance and control of powered-lift short-haul aircraft. This research forms one part of the joint DOT/NASA STOL Operating Systems Experiments Program.

*This paper is based on work done for the Aircraft Guidance and Navigation Branch, Flight Systems Research Division, Ames Research Center, under Contract NAS2-8973.

[†]A versatile digital navigation, guidance, and control system developed by Ames Research Center for conducting experiments with advanced STOL aircraft.

The objective of the joint DOT/NASA STOL Operating Systems Experiments Program is to provide data to aid the design of terminal area guidance, navigation and control systems and the definition of operational procedures for powered-lift and light wing loading short-haul aircraft under IFR. As a first step in this program experimental digital automatic and flight director guidance and control systems have been developed for the NASA Augmentor Wing powered-lift short-haul aircraft by Sperry Flight Systems (under NASA contract). This system, called STOLAND, is based on the application of current CTOL system techniques and displays to the experimental short-haul aircraft (Ref. 1).

Two of the primary displays used in the system are an Electronic Attitude Director Indicator (EADI) and a standard Horizontal Situation Indicator (HSI). In addition, this digital system has a computer driven, cathode ray display called the Multifunction Display, or MFD, which displays the aircraft position and predicted motion on a moving map of the area. Also displayed are other status data including heading, altitude, raw navaid data and reference flight paths. Annotated illustrations of each of the primary displays are shown in Figs. 1-3, and a view of the pilot's instrument panel in the STOLAND simulator used for this investigation is shown in Fig. 4.

This paper presents a preliminary review of some of the comparative measurements and pilot opinions from a flight simulation in February and March 1976. The experimental program for comparing the MFD and HSI within the context of the whole cockpit will be summarized first. Then, after presenting some of the key results, we shall recapitulate our tentative findings in the concluding section of this paper.

EXPERIMENTAL PROGRAM

The experimental investigation was conducted on the NASA/Ames fixed-base STOLAND simulation facility. This facility includes: a) a fully instrumented cockpit; b) a six-degree-of-freedom C-2A Augmentor Wing Jet STOL Research Aircraft/environment/navigation simulation program implemented on an EAI 8400 digital computer; and c) a complete STOLAND digital avionics system.

If the display content has been suited to the task, the display format and symbology will usually be crucial only if the pilot is at a saturated level of workload in a realistic flight simulation or in actual flight. Consequently, we attempted to emphasize the realistic air route navigation, guidance and control environment for short-haul aircraft in preparing the experimental design summarized in Table 1. Three classes of independent variables are shown in the table. The level of pilot involvement in guidance and control tasks is divided between two independent classes, one of which we have called "technique," i.e., either manual or automatic, and the other of which we have called "the level of display," i.e., either situation (raw data only) or flight director and situation on the EADI with the HSI and MFD the obvious independent display variables for comparison.

The flight phases of interest in this experiment were fourfold: a) enroute within 56 km (30 nm) of Crows Landing, ALF, Colusa County, California; b) the

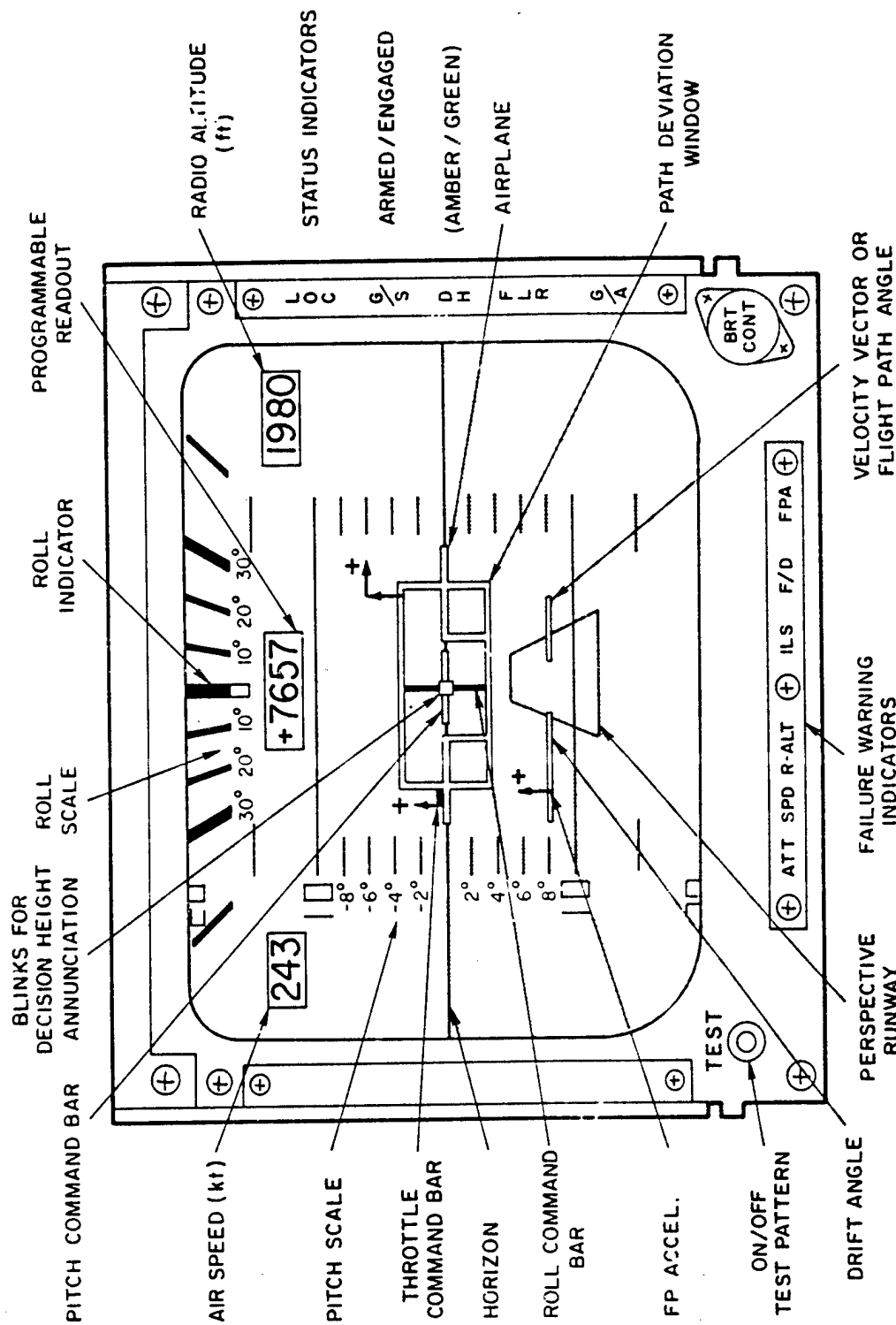


Figure 1. Electronic Attitude Director Indicator (EADI)

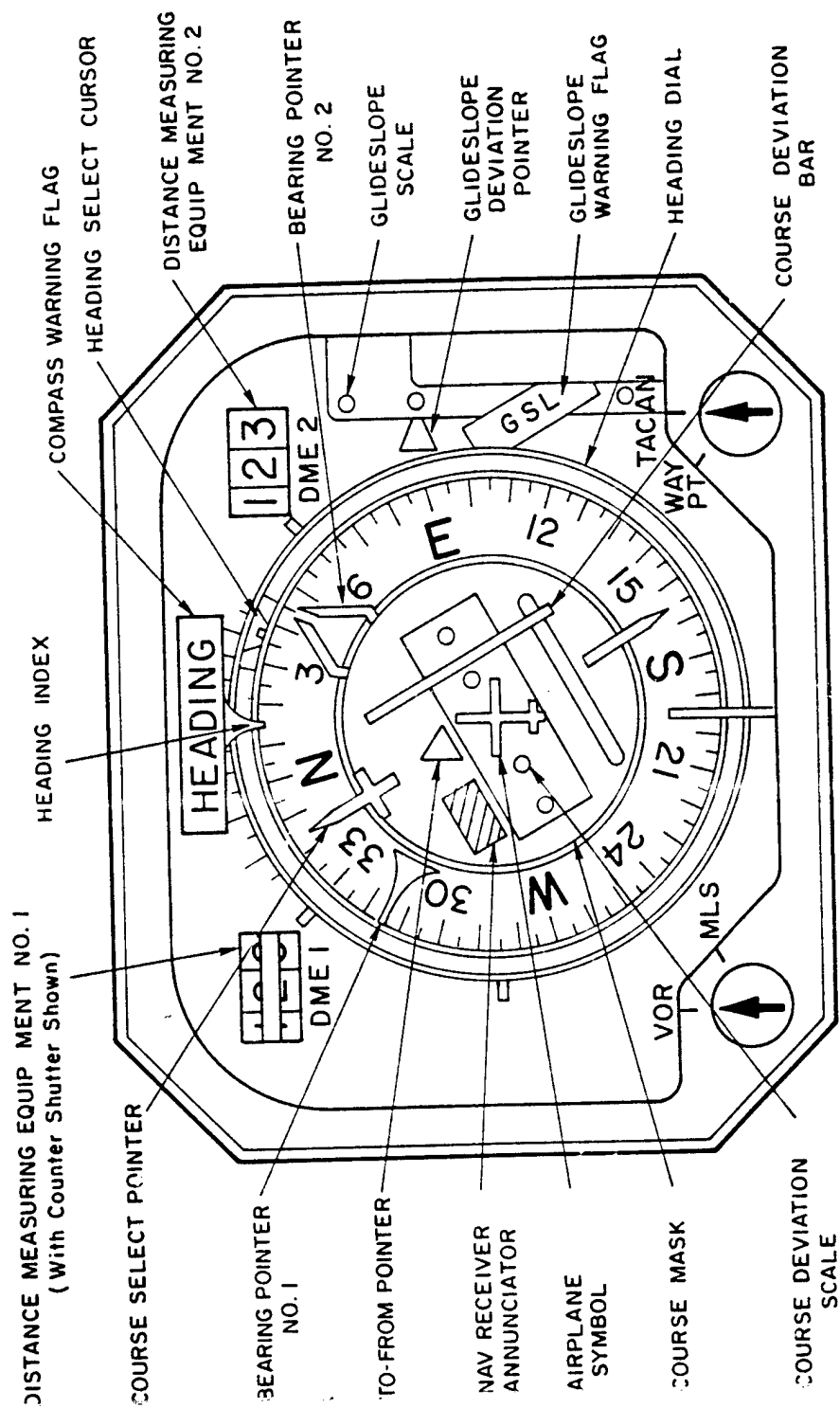


Figure 2. Horizontal Situation Indicator (HSI)

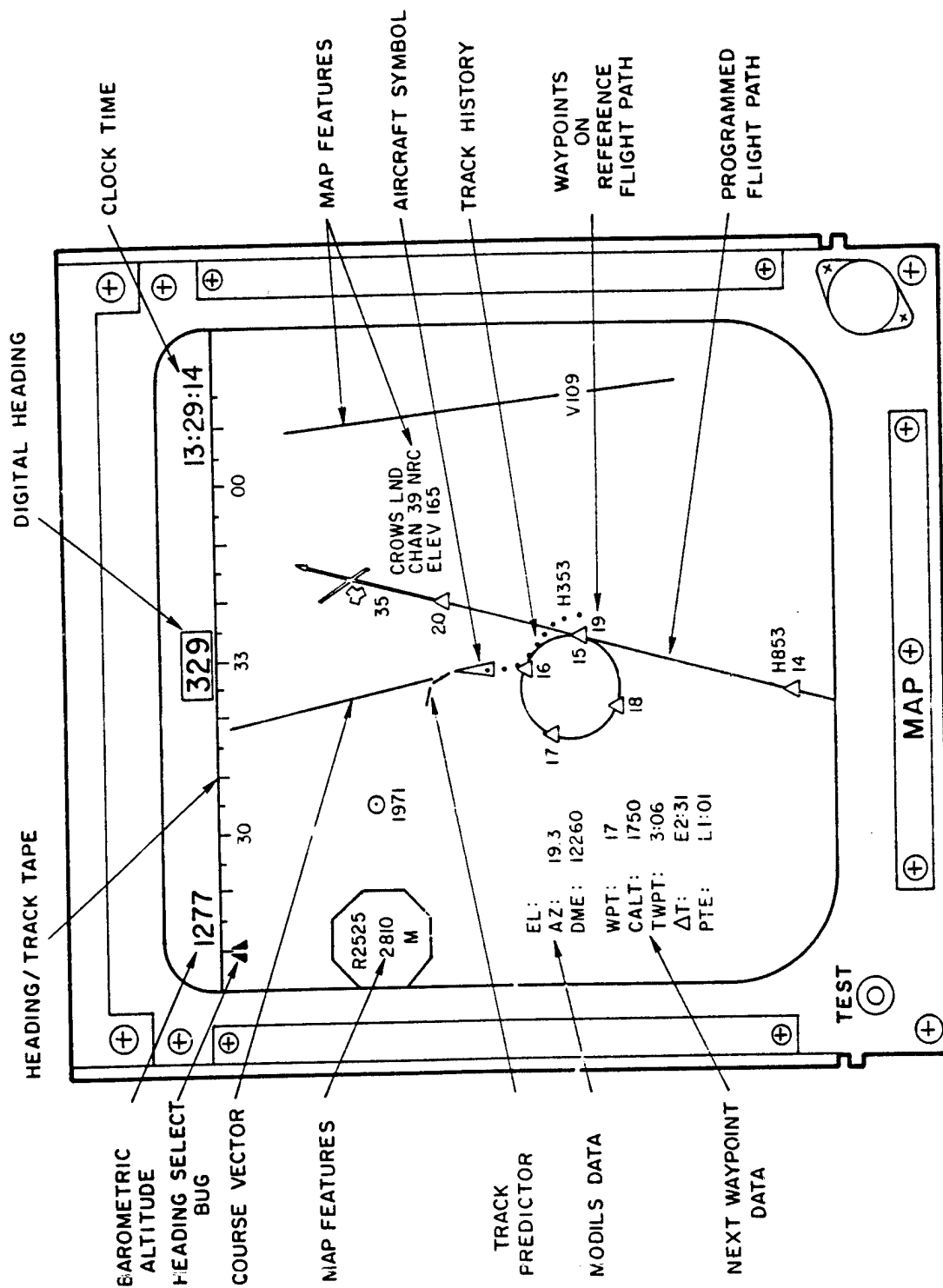


Figure 3. Multifunction Display (MFD)

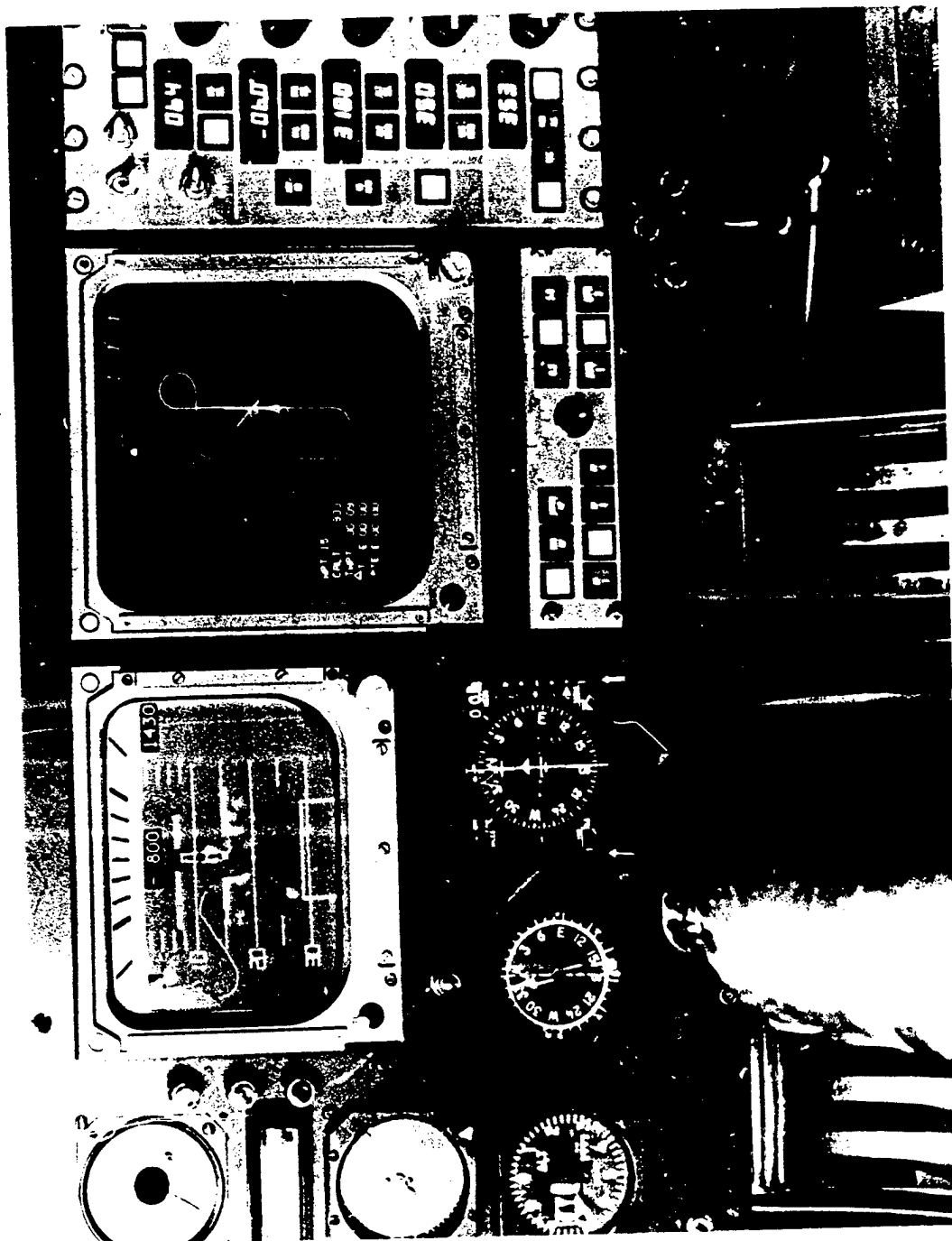


Figure 4. STOLAND Simulator Instrument Panel

TABLE 1. SUMMARY OF EXPERIMENTAL DESIGN

PILOT- ING TECH- NIQUE	NAVIGATION AND CONTROL WORKLOAD	LEVEL OF DISPLAY				
		SITUATION (RAW DATA)		FLIGHT DIRECTOR AND SITUATION ON EADI		
		HSI	MFD	HSI	MFD	BOTH HSI AND MFD
Manual	Tracking a STAR sequence	✓XSCC* Highest Workload	✓ XSCC*	✓	✓	✓ EPR†
	Select different radio nav aids en- route for STOLAND and maintain geo- graphic orientation	✓	✓	✓	✓	✓ EPR†
Auto- matic	Tracking a STAR sequence				Lowest Workload	

10 cells × 2 replications × 5 pilots = 100 runs

*XSCC = Measurement of excess control capacity with
cross-coupled secondary control task

†EPR = Measurement of eye-point-of-regard

Dependent Variables (i.e., Measurements)

a. Flight plan performance errors:

- Airspeed error with respect to commanded flight profile
- Lateral distance error with respect to commanded course
- Altitude or glide slope displacement error
- Elapsed time between waypoints in flight plan

b. Other aircraft motion and control variables (e.g., pitch and roll attitudes, pitch and roll rates, heading, turn rate, airspeed, inertial velocity, angles of attack and sideslip, course and path angles [or ground and vertical velocities], translational accelerations]

c. Eye-point-of-regard in azimuth and elevation

d. Subjective display ratings (e.g., controllability-and-precision, status utility, clutter, attentional demand)

e. Excess control capacity

f. Caution advisory response latency (from a light-cancelling task designed to measure the pilot's simple reaction time)

terminal area itself; c) the landing approach; and d) the go-around involving a holding pattern after a missed approach. We included the four flight phases within a class of independent variables representing the level of pilot involvement in air navigation tasks, i.e., whether the pilot is just following a standard terminal arrival route (STAR) assigned initially by a controller or whether the pilot is additionally involved in reselecting radio nav aids enroute for STOLAND, both during arrival and after executing a missed approach.

While using the automatic mode of STOLAND during initial training sessions, we discovered that automatic operation is so devoid of pilot workload that a critical comparison of the HSI and MFD cannot be made, because the pilot is not even saturated with monitoring tasks. Therefore, we refocused the experimental design on only the manual piloting technique and deferred investigation of the automatic technique.

We have indicated in Table 1 the cells in the experimental design which were most relevant by check marks. We have also indicated the cells which involved the highest and lowest workloads and the two cells which were most amenable to eye-point-of-regard comparison.

Pilot workload is high to begin with when flying the C-3A Augmentor Wing manually with combinations of powered and aerodynamic lift. Since the several STAR's involved holding patterns and curved paths as well as straight segments, reliance on the HSI (and EADI) without the MFD placed the highest workload demand on the pilot, because he had to keep track of his position mentally with the aid of his enroute and terminal area charts as he progressed along the assigned STAR, reselected radio nav aids, and executed missed approaches, go-arounds and holding patterns.

Since the pilot will scan to and fixate on instruments which display redundant information, there is a danger in presenting both the HSI and MFD when the pilot is required to fly with only raw situation data. Having both horizontal displays may actually increase his scanning workload unnaturally when he is already saturated or oversaturated. Therefore, we covered the horizontal display which was not being evaluated in eight cells of Table 1, because the pilot will scan to instruments which display no information or which are temporarily inactive, if given the opportunity. However, both the HSI and MFD were uncovered and presented to the pilot simultaneously in the two cells in the extreme right column of Table 1 when the pilot was using the flight director and situation on the EADI to fly manually. We expected that any outstanding bias in the partitioning of the eye-point-of-regard distribution between the HSI and MFD might afford a measure of pilot preference for or confidence in monitoring the horizontal situation.

The displacement "window" on the EADI was deleted when runs were made to test the HSI alone, since the HSI presents lateral and vertical deviation anyway. The displacement scaling of one "half-window" on the EADI was consistent with the displacement scaling of one dot on the HSI, viz., 381 m (1250 ft) laterally and 30.48 m (100 ft) vertically.

In this connection, we should emphasize that the content of the MFD and HSI are not strictly equivalent, because no waypoint numbers appear on the HSI

and no heading scale appears on the MFD, if the pilot elects the north-up orientation. Furthermore, the format of the altitude presentation on the MFD is purely numerical, and is more unsuitable for tracking than even the counter-pointer altimeter, let alone the vertical deviation indicator (VDI) on the HSI. However, it is as unconventional for the pilots to use the VDI anywhere except on the glide slope as it is to use the displacement "window" on the EADI. Therefore, we may, insofar as the tracking control aspects of this experiment are concerned, be comparing the EADI (supported by the MFD) with the HSI (supported by the EADI without displacement information). Notwithstanding, insofar as the geographic orientation aspects of the experiment are concerned, we are comparing the HSI [supported by an area navigation (RNAV) chart and approach plate] with the MFD, which presents a horizontal moving map of the same RNAV and approach chart.

Heading was provided on the EADI's programmable display during the experiment. We recommended to each pilot that the MFD be used in the course- or heading-up orientation for consistency with the HSI and because the heading tape on the MFD appears only when the course- or heading-up orientation is selected. However, the RMI was always available to present a compass rose when the HSI was covered in the event that a pilot elected to keep the MFD north-up because he experienced disorientation with the revolving map display in turns. The choice of map scale on the MFD was left to the pilot; however, he was instructed that the STAR waypoint numbers would appear only if the 1.5 or 0.5 mn/in. scales were selected.

A steady wind speed of 20 kt from the east or west as required by the flight plan to produce a prevailing tailwind enroute was used throughout the experiment to increase workload, and the wind velocity dispersion was between 3 and 4 ft/sec.

Also listed in Table 1 is the minimum number of 100 runs required for two replications of 10 cells counterbalanced for order effects with 5 pilots. (Over 160 runs were made.) Pilots 1 and 6 are research pilots; Pilots 3 and 4 are commercial airline pilots; Pilot 5 is a general aviation instrument instructor and engineering pilot; and Pilot 2 remained on reserve for this experiment and did not have to participate.

Below the table of independent variables and cells there appears a list summarizing the dependent variables, that is, the measurements which we made. All measurements are self-evident except perhaps "excess control capacity" which is proportional to the value of the spiral divergence required to load the pilot to the point of saturation with control tasks while satisfying primary task performance with respect to a norm or error criterion. Excess control capacity is measured by increasing the spiral divergence until a stationary value is reached by the cross-coupled adaptive regulator of the divergence in balance with the performance error criterion. The regulated average or stationary value of the spiral divergence may be normalized by its critical limit of controllability for each pilot to form a fraction which represents his excess control capacity with respect to the primary task (Ref. 2). To the pilot flying the aircraft the increased spiral divergence seems like a malfunction in lateral stability augmentation, so the measurement can be made while the flight simulation retains high face validity.

Four simple pilot rating scales for use in research on and evaluation of manual control displays were derived and used in the pilot experiments reported in Ref. 3 and were well-suited to the present investigation. The scales shown later in Table 6 are of interval-scale quality and will permit averaging and other standard parametric statistical analyses. The use of four trait categories (task controllability and precision, status utility, clutter, and attentional demand) helped to separate subjective identification of these often confounded effects. Rating forms for the HSI and MFD were filled out by each pilot in the cockpit at the conclusion of a simulated flight.

RESULTS OF THE EXPERIMENT

Although we shall continue to label the displays being compared as "HSI" and "MFD" for conciseness in presenting the results where one or the other horizontal display was uncovered, the reader should clearly understand that "HSI" means "HSI, EADI (without the displacement window) and other instruments" and that "MFD" means "EADI, MFD and other instruments." By design, the HSI and MFD are being compared within the context of the whole STOLAND display and control arrangement in the simulator cockpit.

We shall now turn to present the several forms of comparative results of the experiment under the following subordinate topical headings:

- Blunders
- Tracking errors
- Excess control capacity
- Excess monitoring capacity
- Pilot opinion ratings
- Eye-point-of-regard

Blunders

About 160 simulated flights, each lasting from 10 to 25 minutes in time, were conducted among four standard terminal arrival routes (STAR's). The most dramatic results are the 20 "blunders" partitioned in Tables 2 and 3. Table 2 partitions the 9 blunders which occurred in the first phase of the experiment in February while the pilots were primarily tracking STAR's as reference flight paths. Table 3 partitions the remaining 11 blunders which occurred during terminal area and enroute flight with emphasis on geographic orientation (as well as tracking) in the second phase of the experiment in March involving missed approaches, go-arounds and holding patterns with three different radio nav aids. The types of blunders identified include loss of geographic orientation, loss of altitude awareness, and loss of roll attitude control as well as several others.

Five blunders involved the HSI and 4 the MFD, while tracking reference flight paths exclusively (Table 2). However, 8 blunders involved the HSI, 2 the MFD, and 1 both displays, during terminal area and enroute flight with emphasis on geographic orientation (Table 3). The flight director was (or should have been) in use during 11 of the 20 runs wherein blunders occurred. Since 7 of these 11 blunders were also associated with the HSI in Table 3, the combination of using the HSI for orientation with the flight director for tracking while selecting different radio nav aids for guidance seemed to conspire to produce the most blunders. There were no blunders involving the MFD and flight director in Table 3 and only three in Table 2. Therefore, we would conclude

TABLE 2

DISTRIBUTION OF BLUNDERS WHILE TRACKING REFERENCE FLIGHT PATHS

BLUNDERS	RAW DATA		FLIGHT DIRECTOR	
	HSI	MFD	HSI	MFD
Loss of geographic orientation	2	None	None	None
Loss of altitude awareness	1	None	None	1
Loss of roll attitude control	1	1	None	2
Other crashes	1 GPIP ^a	None	None	None

^aGlide path intercept point

TABLE 3

DISTRIBUTIONS OF BLUNDERS DURING TERMINAL AREA AND ENROUTE FLIGHT
WITH EMPHASIS ON GEOGRAPHIC ORIENTATION

BLUNDERS	RAW DATA		FLIGHT DIRECTOR ^b	BOTH HSI AND MFD WITH FD
	HSI	MFD	HSI	
Loss of geographic orientation	None	None	3	None
Loss of altitude awareness	None	2 (Crash and missed capture)	1	None
Loss of roll attitude control	None	None	2	1
"Copilot error"	None	None	1	None
Experimenter's error	1	None	None	None

^bThere were none with MFD and Flight Director.

on the basis of the blunder distribution alone from the simulation that the MFD seems to offer a worthwhile improvement in safety, since 13 of 20 blunders involved runs wherein the MFD was not available to the pilot

Tracking Errors

The least dramatic results are to be found among the flight plan tracking errors in three dimensions (lateral and vertical displacement and airspeed) and the related variations in aircraft motions. As one would expect, the flight director provided for more precise tracking of the assigned altitude and the glide slope than otherwise. However, there was no consistent evidence of differences between tracking errors with the HSI versus the MFD even with only raw data. Yet, as we mentioned in beginning the discussion of results, the "MFD" implies the use of the integrated EADI as the tracking display, and occasionally better altitude-keeping performance appeared with the "MFD" than with the "HSI."

Excess Control Capacity

The measurement of excess control capacity was provided by the average cross-coupled adaptive spiral divergence in selected runs with either the HSI or the MFD. The null hypothesis of equality between mean values of excess control capacity within comparable pairs of waypoint groups with either display arrangement was tested for significant differences. The results of these tests are listed in Table 4 by pilot and flight plan. The column heading "neither"

TABLE 4

NUMBER OF COMPARABLE PAIRS OF WAYPOINT GROUPS FOR WHICH ONE OR THE OTHER DISPLAY ARRANGEMENT EXHIBITED SIGNIFICANTLY GREATER AVERAGE EXCESS CONTROL CAPACITY AT THE 0.05 LEVEL^a

<u>PILOT</u>	<u>FLIGHT PLAN</u>	<u>HSI/EADI</u>	<u>NEITHER</u>	<u>EADI/MFD</u>
1	2	1	4	1
3	2	3	4	1
3	3	1	1	4
4	2	5	1	6
5	2	<u>1</u>	<u>1</u>	<u>4</u>
Totals		11	11	16

^aThe null hypothesis is "neither." The probability of rejecting the null hypothesis when it is true is 0.05. Behrens', Scheffe's, and Tukey's tests produced consistent results.

identifies the number of comparable pairs of waypoint groups for which the null hypothesis was accepted. The probability of rejecting the null hypothesis when it is true is 0.05. Behrens', Scheffe's, and Tukey's tests produced consistent results under the ergodic hypothesis, because the number of samples available within each waypoint group was on the order of several hundred or more.

The column headings "HSI" or "MFD" identify the numbers of comparable pairs of waypoint groups for which the null hypothesis was rejected, i.e., for which one or the other display arrangement exhibited significantly greater average excess control capacity at the 0.05 level. The totals show that the null hypothesis was rejected for 27 of 38 pairs at the 0.05 level. Of these 27 pairs, the "MFD" exhibited greater average excess control capacity for 16, and the "HSI" greater for 11 pairs. In the individual case of Pilot 3 tracking Flight Plan 2 involving only a curved approach, the partition is in favor of the "HSI," a result which was consistent with that pilot's own appraisal of that flight plan. However, the partition for Pilot 3 with Flight Plan 3, involving a missed approach and holding pattern is in favor of the MFD.

Excess Monitoring Capacity

The measurement of excess monitoring capacity was inversely proportional to the average caution advisory response time. The null hypothesis of equality between mean response times within comparable pairs of runs with either display arrangement was tested for significant differences after a correction for the skewness of the response time distribution was made. The results of these tests are listed in Table 5 by pilot. The column heading "neither" identifies

TABLE 5

NUMBER OF COMPARABLE PAIRS OF RUNS FOR WHICH ONE OR THE OTHER
DISPLAY ARRANGEMENT EXHIBITED SIGNIFICANTLY GREATER AVERAGE
EXCESS MONITORING CAPACITY AT THE 0.05 LEVEL^a

<u>PILOT</u>	<u>HSI/EADI</u>	<u>NEITHER</u>	<u>EADI/MFD</u>
1	1	4	1
3	0	7	2
4	0	9	2
5	1	4	1
6	<u>1</u>	<u>1</u>	<u>4</u>
Totals	3	25	10

^aThe null hypothesis is "neither." The probability of rejecting the null hypothesis when it is true is 0.05. Behrens', Scheffe's, and Tukey's tests produced consistent results.

the number of comparable pairs of runs for which the null hypothesis was accepted. The probability of rejecting the null hypothesis when it is true is 0.05. Again Behrens', Scheffe's, and Tukey's tests produced consistent results, because there were usually at least eleven samples in the ensemble for each run. The column headings "HSI" or "MFD" identify the numbers of comparable pairs of runs for which the null hypothesis was rejected, i.e., for which one or the other display arrangement exhibited significantly greater average excess monitoring capacity at the 0.05 level. The totals show that the null hypothesis was rejected for 13 of 38 pairs at the 0.05 level. Of these 13 pairs, the "MFD" exhibited greater average excess control capacity for 10, and the "HSI" greater for 3 pairs.

Pilot Opinion Ratings

Tables 6a and 6b present summaries of the subjective opinion ratings of the HSI and MFD by each of Pilots 1, 3, 4, and 6 during the second phase of the experiment in March emphasizing geographic orientation as well as tracking. A comparison of Tables 6a and 6b shows a slightly less favorable central tendency in the ratings of the task controllability and precision when using the HSI, whereas the ratings are more uniformly distributed over four descriptive phrases when using the MFD. Ratings of task controllability and precision with the flight director in use are uniformly distributed over four descriptive phrases when using either the HSI or MFD. Comparison of the ratings for utility of status information between the HSI and MFD shows more favorable ratings for the MFD and a markedly unfavorably skewed distribution of ratings for the HSI which exhibits a mode beside the descriptive phrases: (S4) "inadequate number of states...." Comparison of the ratings for clutter shows few differences in the tendency of both groups of ratings to centralize beside the descriptive phrase: (K3) "some clutter." Only one rating of the MFD was more unfavorable than K3. Comparison of the ratings for display attentional workload shows a more favorable central tendency beside the descriptive phrase: (D3) "mildly demanding" for the MFD, whereas the distribution of ratings for the HSI is unfavorably skewed with a mode beside the descriptive phrase: (D4) "quite demanding."

During the tracking of reference flight paths in the first phase of the experiment in February, slightly different central tendencies in some of the ratings by Pilots 3, 4, and 5 were observed. For example, the ratings for utility of status information on the HSI were bimodally distributed between (S2) "many of desired states presented" and (S4) "inadequate number of states." The rationale was apparent from the accompanying commentary, viz., that the HSI is quite adequate for tracking rectilinear courses whether inclined or not, whereas the HSI is deficient for tracking curved courses in the presence of wind. Ratings of the utility of status information on the MFD were skewed favorably with a mode beside S2. A few more "quite cluttered" (K4) ratings of the MFD were received, although the central tendency on both HSI and MFD remained beside K3. Ratings of attentional workload while tracking with either HSI or MFD were centered on (S4) "quite demanding," although relatively more (D5) "completely demanding" ratings of the HSI were given.

TABLE 6a

**SUMMARY OF 3 RATINGS OF HSI BY EACH OF PILOTS 1, 3, 4, 6 DURING SECOND
PHASE OF EXPERIMENT EMPHASIZING GEOGRAPHIC ORIENTATION**

PILOT OPINION RATING SCALES

RATING SCALE FOR UTILITY OF STATUS INFORMATION

CRITERIA	DESCRIPTIVE PHRASE	RATING
Usefulness of the information supplied, on the specific display unit, on the vehicle status — especially the relevant flight path vector status, such as: altitude, speed, heading attitude, path error, etc.	All desired states presented with adequate resolution and readability	S1
	Many of desired states presented, with a few deficiencies in scaling, resolution, or readability	S2
	Some desired states presented, and/or some problems with scaling, resolution, or readability	S3
	Inadequate number of states, or serious deficiencies in scaling, resolution, or readability	S4
Useful with respect to the mission phase, task criteria, and operator's sense of vehicle safety.	No direct status information or unusable	S5

RATING SCALE FOR CLUTTER

CRITERIA	DESCRIPTIVE PHRASE	RATING
Degree of subjective symbol-clutter on specified display unit	Completely uncluttered — e.g., only one pair of elements	K1
	Mostly uncluttered — no competing or distracting elements	K2
	Some clutter — multiple elements competing for attention	K3
	Quite cluttered — difficult to keep track of desired quantities among competitors	K4
	Completely cluttered — nearly impossible to tell desired elements or quantities due to competing elements	K5

RATING SCALE FOR TASK CONTROLLABILITY AND PRECISION

CATEGORY	DESCRIPTIVE PHRASE	RATING
CONTROLLABLE	Very easy to control, with good precision	C1
	Easy to control, with fair precision	C2
Yes	Controllable, with inadequate precision	C3
	Marginally controllable	C4
No	Uncontrollable	C5

RATING SCALE FOR DISPLAY ATTENTIONAL WORKLOAD

CRITERIA	DESCRIPTIVE PHRASE	RATING
Demands on the operator attention, skill, or effort	Completely undemanding and relaxed	D1
	Mostly undemanding	D2
	Mildly demanding	D3
	Quite demanding	D4
	Completely demanding	D5

* Controllable with difficulty or high workload, but fair precision

⊕ = Raw data. ⊕ = Flight director and situation.

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 6b

SUMMARY OF 3 RATINGS OF MFD BY EACH OF PILOTS 1, 3, 4, 6 DURING SECOND
PHASE OF EXPERIMENT EMPHASIZING GEOGRAPHIC ORIENTATION

PILOT OPINION RATING SCALES

RATING SCALE FOR UTILITY OF STATUS INFORMATION

CRITERIA	DESCRIPTIVE PHRASE	RATING
Usefulness of the information supplied, on the specified display unit, on the whole status — especially the relevant flight path vector status, such as: altitude, speed, heading attitude, path error, etc.	All desired status presented with adequate resolution and readability	S1
	Many of desired status presented, with a few deficiencies in scaling, resolution, or readability	S2
	Some desired status presented, and/or some problems with scaling, resolution, or readability	S3
	Inadequate number of status, or serious deficiencies in scaling, resolution, or readability	S4
Partial with respect to the mission phase, task criteria, and operator's sense of vehicle safety.	No direct status information or unusable	S5

RATING SCALE FOR CLUTTER

CRITERIA	DESCRIPTIVE PHRASE	RATING
Degree of subjective symbol-clutter on background specified display unit	Completely uncluttered — e.g., only one pair of elements	K1
	Mostly uncluttered — no annoying or distracting elements	K2
	Some clutter — multiple elements competing for attention	K3
	Quite cluttered — difficult to keep track of desired quantification among competitors	K4
	Completely cluttered — nearly impossible to tell desired elements or quantifies due to competing elements	K5

RATING SCALE FOR TASK CONTROLLABILITY AND PRECISION

CATEGORY	DESCRIPTIVE PHRASE	RATING
CONTROLLABLE		
Yes	Very easy to control, with good precision	C1
	Easy to control, with fair precision	C2
	Controllable, with inadequate precision	C3
No	Marginally controllable	C4
	Uncontrollable	C5

RATING SCALE FOR DISPLAY / ATTENTIONAL WORKLOAD

CRITERIA	DESCRIPTIVE PHRASE	RATING
Depends on the operator attention, effort, or effort	Completely un-demanding and relaxed	D1
	Mostly undemanding	D2
	Mildly demanding	D3
	Quite demanding	D4
	Completely demanding	D5

* Controllable with difficulty or high workload, but fair precision

* Flight director and situation.

Eye-Point-of-Regard (EPR)

This experiment has resulted in the acquisition of our largest archive of high quality data to date with the STI Eye-Point-of-Regard System Model EPR-2. However, it will be possible to review only a small sample of the EPR data here. The EPR data acquisition was confined to runs wherein both HSI and MFD (as well as all other active displays and controls in the cockpit) were available to the pilot in accord with the experimental plan in Table 1. The reduced data to be presented are from the go-around phase of four runs by two pilots. Figures 5-8 show EPR dwell fractions, look fractions, and transition link fractions for the pilots using either (a) only raw situation data or (b) the flight director with situation data during the go-around phase to a holding fix following a missed approach.

The dwell fraction is merely the relative dwell time-weighted look fraction, and the look fraction is simply the relative number of fixations on each instrument or display. The look fraction represents the ensemble probability of fixation and the dwell fraction, the temporal probability of fixation. The bidirectional link fraction is the relative number of scan transitions in both directions between each pair of instruments or displays. In rare cases, transitions occurred in only one direction noted by an arrow. The sums of each type of fraction may not equal exactly unity, because of round-off errors in the listed values.

Comparison of Figs. 5 and 6 shows that Pilot 3 used the HSI more when the flight director was off and the MFD more when the flight director was on. This is exemplified by all three types of fractions. Pilot 3 was consistent in this dichotomy. There are probably two underlying reasons:

- a. When using raw situation data, it may be easier to close the heading control loop with the more familiar HSI, and
- b. When using the flight director on the EADI, it may be easier to monitor the aircraft's heading and geographic position simultaneously using the moving map display on the MFD.

Direct crosschecks between the HSI and the MFD, although rare, do appear in one direction from the HSI to the MFD in the results in Figs. 5 and 7 with only raw data.

Pilot 1 (Figs. 7 and 8) used the HSI hardly at all during the go-around phase to a holding fix with raw data and not at all during the go-around with the flight director. The absence of any transitions to the altimeter in Fig. 8 may be because barometric altitude was available on the MFD, although we have not yet analyzed subsidiary transitions within the face of the MFD.

Pilot 1, a research pilot, was much more familiar with the EADI and MFD than Pilot 3, a commercial airline pilot. The familiarity and confidence of Pilot 1 in using the MFD is evident in comparing his eye-point-of-regard distribution with that for Pilot 3, in spite of the extensive training time provided for the commercial pilots who participated in the experiment.

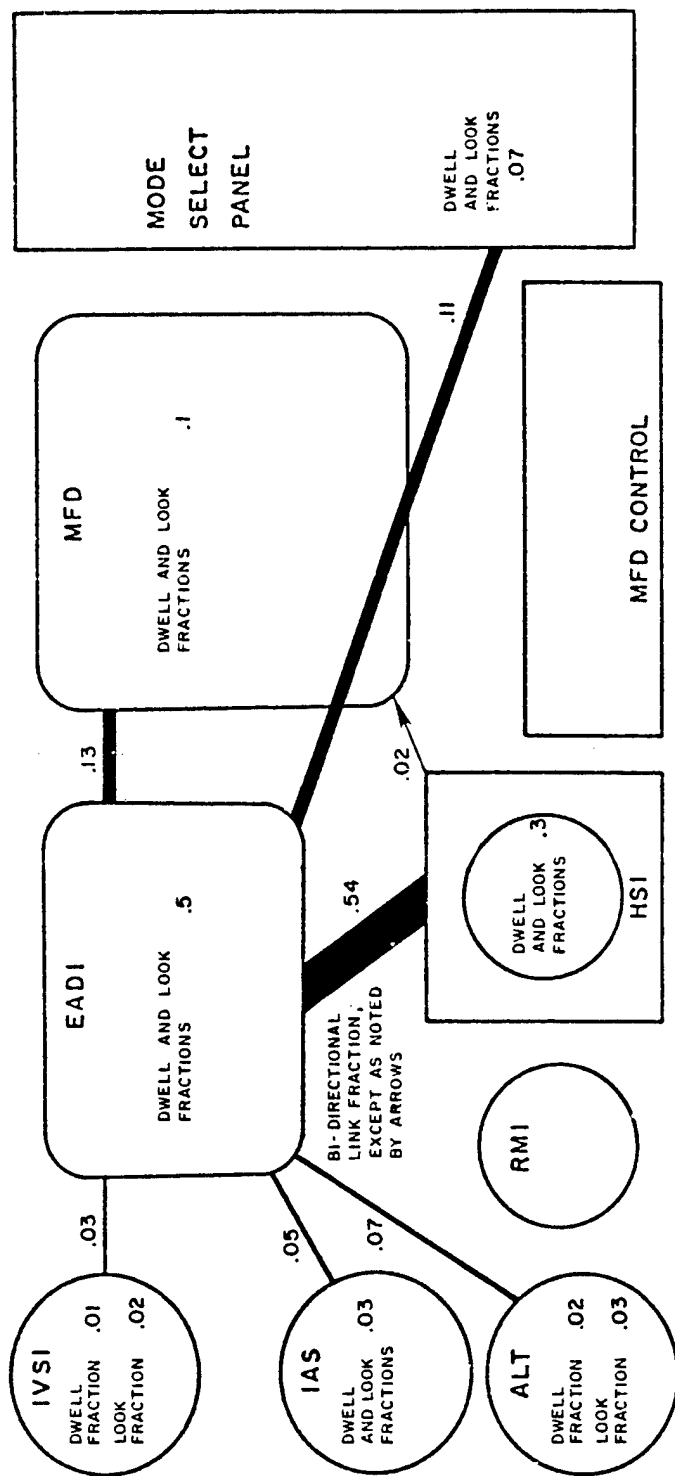


Figure 5. Eye-Point-of-Regard Dwell Fractions, Look Fractions, and Transition Link Fractions Averaged for Pilot 3 Using Raw Data During the Go-Around Phase to a Holding Fix Following a Missed Approach (Waypoints 22-25 of Run 148 with Flight Plan 3)

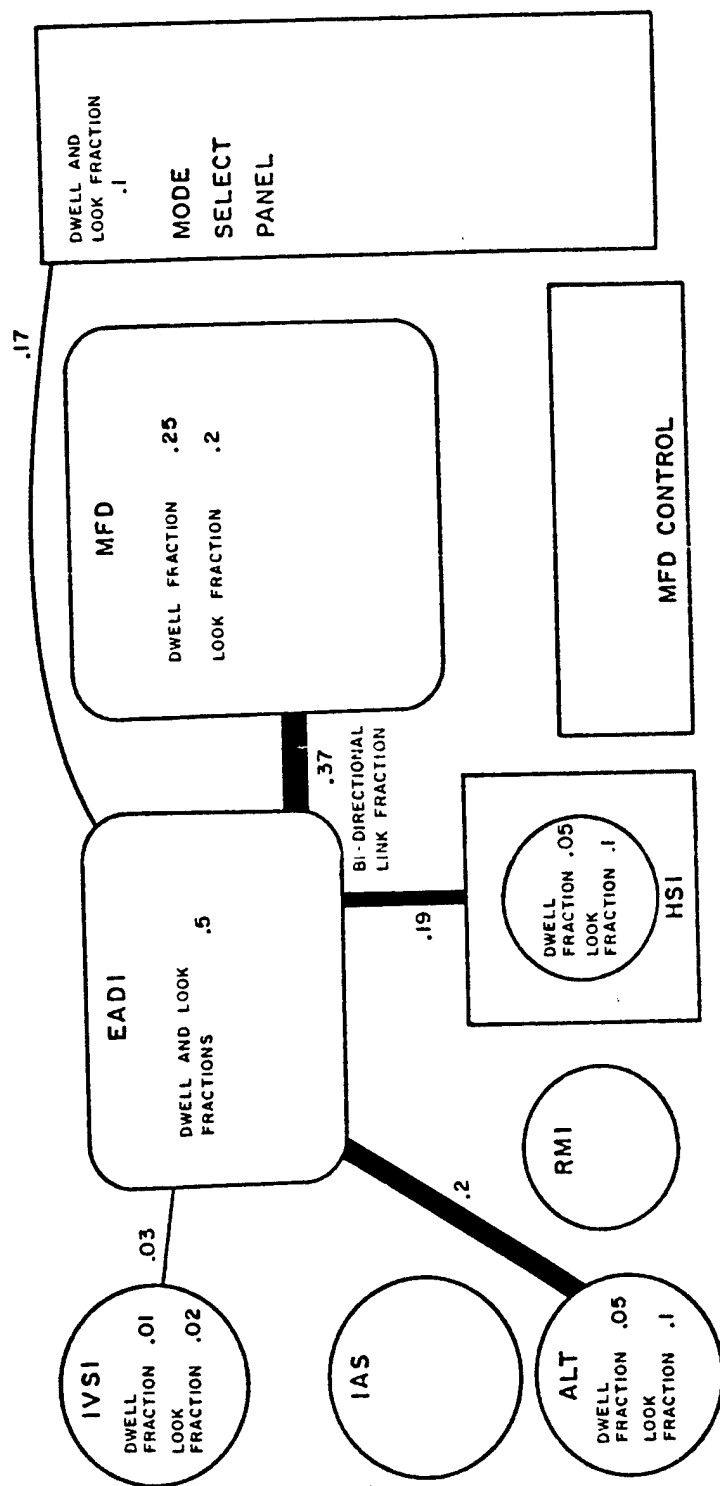


Figure 6. Eye-Point-of-Regard Dwell Fractions, Look Fractions, and Transition Link Fractions Averaged for Pilot 3 Using the Flight Director During the Go-Around Phase to a Holding Fix Following a Missed Approach (Waypoints 22-25 of Run 149 with Flight Plan 3)

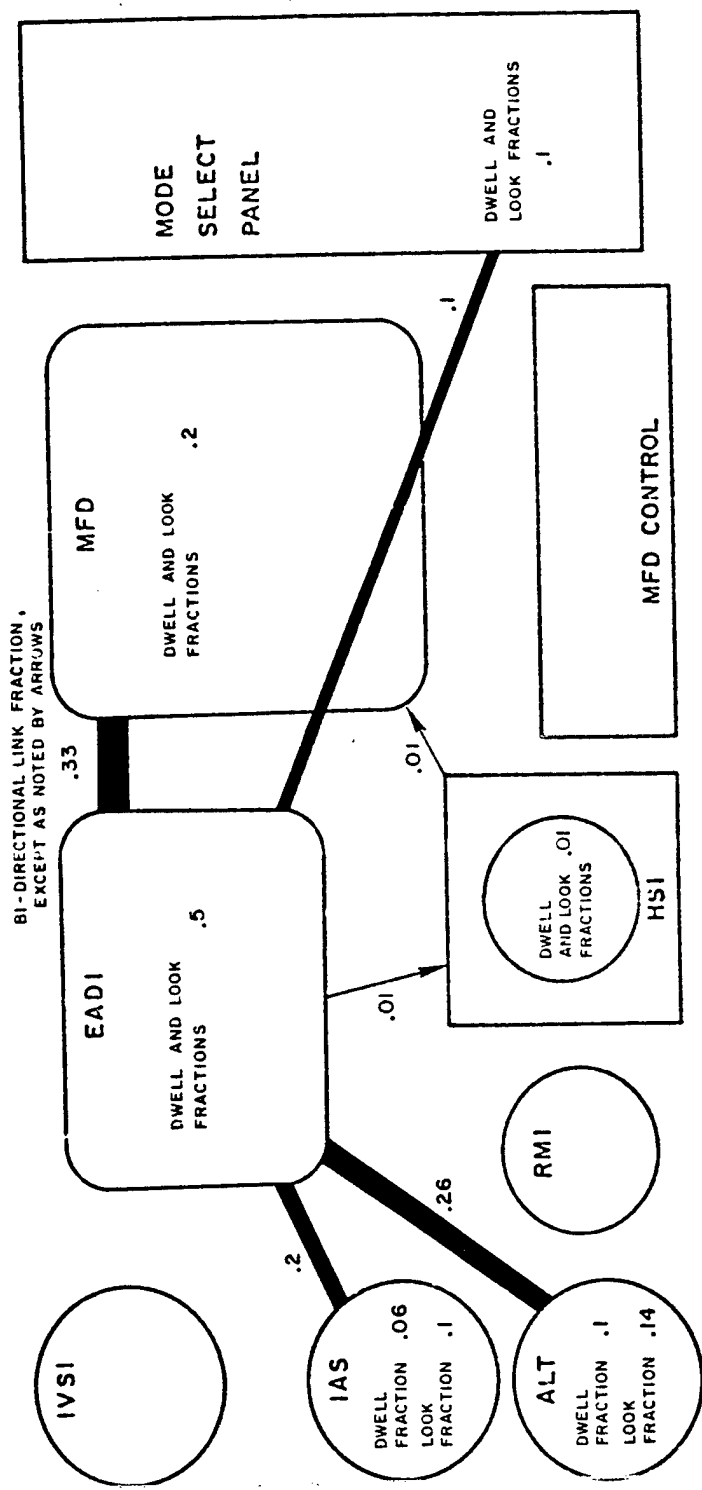


Figure 7. Eye-Point-of-Regard Dwell Fractions, Look Fractions, and Transition Link Fractions Averaged for Pilot 1 Using Raw Data During the Go-Around Phase to a Holding Fix Following a Missed Approach (Waypoints 22-25 of Run 15+ with Flight Plan 3)

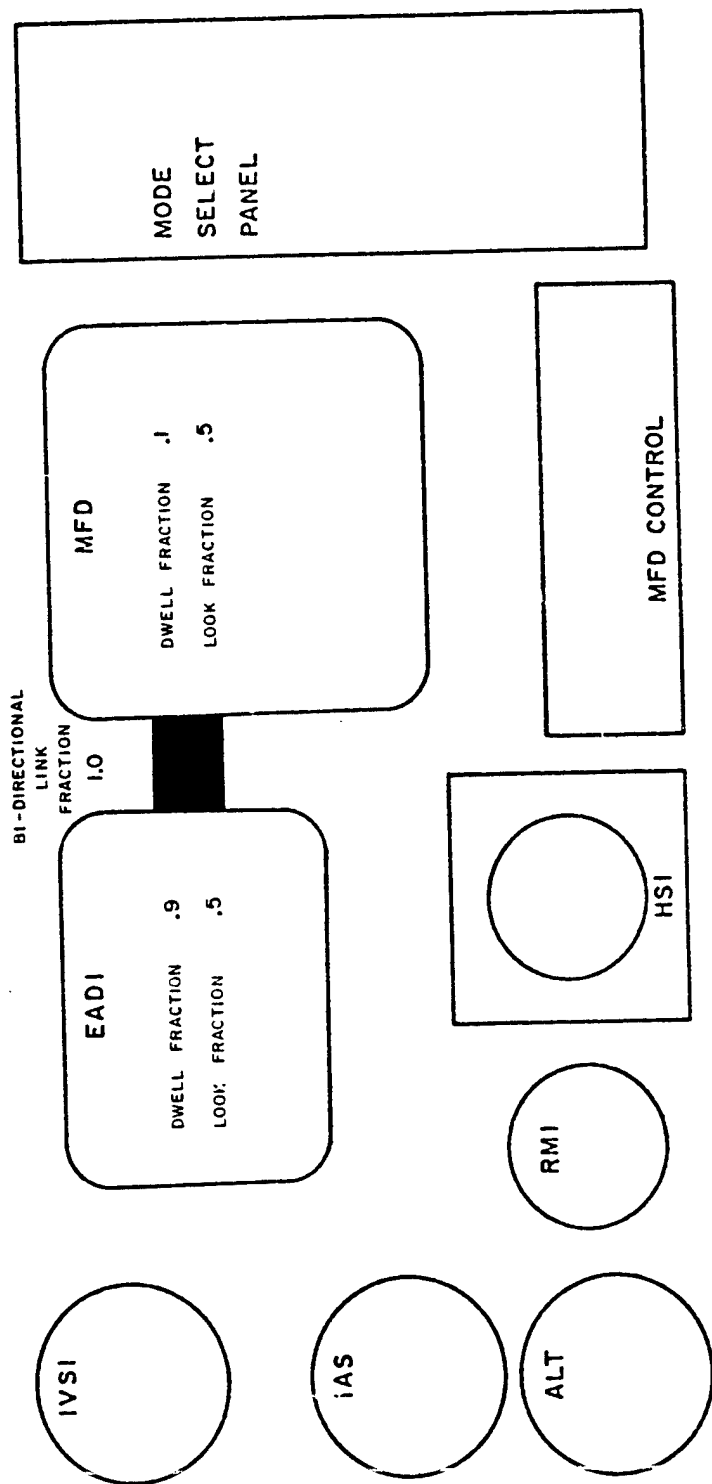


Figure 8. Eye-Point-of-Regard Dwell Fractions, Look Fractions, and Transition Link Fractions Averaged for Pilot 1 Using the Flight Director During the Go-Around Phase to a Holding Fix Following a Missed Approach (Waypoints 22-25 of Run 15, with Flight Plan 3)

Although we must conclude this brief preview of results, we shall, in the summary which follows, provide a concise overview of some of the other trends in the eye-point-of-regard distributions.

SUMMARY OF RESULTS AND CONCLUDING REMARKS

On the basis of the blunder distribution alone from the simulation, the MFD seems to offer a worthwhile improvement in safety, since 13 of 20 blunders among 160 runs involved runs wherein the MFD was not available to the pilot. Six involved runs with the MFD, but not the HSI, and only one involved a run with both.

The flight director provides for more precise tracking of the assigned altitude and the glide slope than otherwise. However, there is no consistent evidence of differences between tracking errors with the HSI versus the MFD among the five practiced pilots who participated in the simulation, although there are instances where altitude-keeping was more precise with the MFD when using only situation data.

The null hypothesis of equality between average excess control capacity within comparable pairs of flight plan waypoint groups using either the HSI or the MFD arrangement was tested for significant differences. The results show that the null hypothesis was rejected for 27 of 38 pairs of comparable waypoint groups at the 0.05 level, where 0.05 is the probability of rejecting the null hypothesis when it is true. Of these 27 rejected pairs, the MFD exhibited greater average excess control capacity for 15, and the HSI greater for 11 pairs.

The null hypothesis of equality between average excess monitoring capacity within comparable pairs of runs using either the HSI or the MFD arrangement was tested for significant differences. The results show that the null hypothesis was rejected for 13 of 38 pairs of comparable runs at the 0.05 level. Of these 13 rejected pairs, the MFD exhibited greater average excess monitoring capacity for 10, and the HSI greater for 3 pairs.

The pilots provided subjective ratings of (a) task controllability and precision, (b) utility of status information, (c) symbol-background clutter on the display, and (d) display attentional workload, each on five-point descriptive scales. Summarized comparisons of all ratings for the MFD and the HSI in each category follow.

Task Controllability and Precision. There is a slightly less favorable central tendency to rate the HSI "controllable, with inadequate precision," in tracking curved paths in the presence of wind, whereas the ratings favor the MFD as "controllable, with fair precision." Ratings with the flight director in use are uniformly distributed over four descriptive phrases from "easily" to "marginally" controllable and exhibit no central tendency with either HSI or MFD.

Utility of Status Information. The MFD received more favorable ratings than the HSI. Ratings of the usefulness of information supplied covered

the three adjectives "adequate": "some": "inadequate" in the ratios 4:3:0 for the MFD and 4:2:5 for the HSI. The bimodality of ratings for the HSI is believed to be associated with the favorable view of the HSI for tracking rectilinear flight paths and the unfavorable view of the HSI for maintaining geographic orientation while tracking curved paths in the presence of wind.

Clutter. Ratings of both the HSI and MFD as having "some clutter" pre-dominate, but there is a slight tendency to rate the MFD less favorably.

Display Attentional Workload. Ratings of the MFD show a central tendency between "mildly" and "quite demanding," whereas ratings of the HSI exhibit a less favorably skewed mode between "quite" and "com-pletely demanding."

The reduced eye-point-of regard (EPR) data for 8 runs among Pilots 1, 3, and 4 reveal that, with few exceptions, there are relatively more looks at and longer fixation dwells on the MFD than the HSI when using raw situation data and especially when using the flight director. This finding is consistent with the respective comments by Pilots 1 and 4 and tends to confirm an expressed preference by Pilot 1 for the MFD in curved path tracking and in negotiating holding patterns. While not definitive in the case of Pilot 4 because of the single flight plan, the EPR data from two runs also suggest a preference by Pilot 4 for the MFD in curved path tracking. In the case of Pilot 3 the EPR data suggest an equitable distribution of looks and dwells between the HSI and MFD throughout the approach with the flight director, but a preference for the MFD during the missed approach, go-around, and throughout the holding pattern in Flight Plan 3. In the case of Pilot 3 using raw data, the EPR measurements offer little basis for inferring a preference between the HSI and MFD, because both horizontal displays are scanned, in turn, from the EADI fairly consistently throughout Flight Plan 3, except during the straight final approach where the HSI receives relatively more looks. Direct crosschecking between the HSI and the MFD is rare, but such scan transitions are pre-ominantly uni-directional from HSI to MFD by Pilots 1, 3, and 4 when using raw situation data.

All of the pilots provided a great number of helpful supporting comments in the course of the experiment. There was a consensus among the pilot comments which acknowledged the excellence of the HSI for tracking rectilinear inclined courses, yet recognized the superiority of the MFD for maintaining confidence in geographical orientation while tracking curved approach courses and establishing holding patterns in the presence of wind. The summary comment by one of the pilots provides a decisive conclusion: "If the MFD in operational aircraft could be shared with the weather display...between the pilots, a moving map display might be provided at no extra cost in panel space. The HSI bearings, DME, and deviations would seem to be an essential backup, even for a moving map. I would...want both HSI and MFD, if the MFD could be shared with the weather radar display....The HSI improves my confidence in the moving map on the MFD."

REFERENCES

1. Neuman, F.; Watson, D. M.; and Bradbury, P.: Operational Description of an Experimental Digital Avionics System for STOL Airplanes. NASA TM X-62,448, December 1975.
2. Clement, W. F.; McRuer, D. T.; and Klein, R. H.: Systematic Manual Control Display Design. Proceedings of the 13th AGARD Guidance and Control Symposium on Guidance and Control Displays. AGARD CP-96, February 1972, pp. 6-1 to 6-10.
3. Clement, W. F.; Allen, R. W.; and Graham, D.: Pilot Experiments for a Theory of Integrated Display Format. JANAIR Report 711107, October 1971 (also TR-183-2, Systems Technology, Inc.).

AIR-TO-GROUND VISUAL DISPLAY SYSTEM

Daniel L. Kugel
Randall V. Gressang, Capt, USAF
Air Force Flight Dynamics Laboratory
Wright-Patterson AFB, Ohio

ABSTRACT

The ability to simulate realistic air-to-ground tactical missions is very important in the evaluation of prototype fighter aircraft, testing of new Air Force weapon systems, and the training of pilots for high performance aircraft. Most aerospace simulators however do not have this capability. The Air Force Flight Dynamics Laboratory at Wright-Patterson AFB, Ohio has recently developed, in a joint effort with the Northrop Corporation, a Large Amplitude Multi-mode Aerospace Research Simulator (LAMARS). The use of this simulator for air-to-ground tactical mission simulation is discussed. The utilization of a state-of-the-art helmet sight system and its role in the development of a head slaved visual display system is explained; the major subsystems are described, with principle emphasis placed on the visual display system, terrain model, and helmet sight system. Engineering data gathered to evaluate this area of interest (AOI) display is included, along with details of the drive logic and a summary of pilot comments evaluating the ability of the visual display system to provide the necessary requirements to perform the air-to-ground mission.

I. INTRODUCTION

The LAMARS engineering simulator is currently in operation at the AF Flight Dynamics Laboratory. This high performance simulator will serve as one of the major engineering tools in the new ground-based simulation facility, the Flight Control Development Laboratory. This facility is the focal point for coordinated studies encompassing pilot/vehicle studies, ground based and in-flight simulation as applied to flight dynamic research, prototype evaluation and weapon system development. The LAMARS will be used in support of programs covering advanced development, system development and integration, and simulation support.

The combination of wide angle visual cues and large amplitude motion cues that the simulator gives to the pilot provides an environment which will evoke proper pilot response characteristics for a large number of mission conditions and pilot control tasks. Knowledge of how a pilot will react to various conditions is required during preliminary design and preflight evaluation of military aircraft designs. Such prior knowledge will identify possible trade-offs between aircraft performance, stability and control, and maneuverability. The engineering flight simulator gives the Air Force pilot the

unique opportunity to significantly influence the design of a flight vehicle long before the final configuration is determined.

Though not its principle mission, the Flight Control Division of the AF Flight Dynamics Laboratory has the engineering knowledge and first hand hardware experience to help make determinations of simulator criteria and requirements. Recently, the Simulator System Program Office of the Aeronautical Systems Division, Wright-Patterson AFB began work on testing present Air Force simulators for the purpose of developing visual requirements for air-to-ground mission simulators. This study was to determine both qualitatively and quantitatively the cues which could be generated visually by the LAMARS simulator at the AF Flight Dynamics Laboratory, Wright-Patterson AFB, Dayton, Ohio, the Advanced Simulator for Undergraduate Pilot Training (ASUPT) at Williams AFB, Chandler, Arizona, and the Simulator for Air-to-Air Combat (SAAC) at Luke AFB, Glendale, Arizona. By evaluating these three methods of visual projection one can better determine which methods will best provide the needed visual cues to perform the air-to-ground tactical mission.

It should be noted that this program was conducted to show what present Air Force simulators could provide in the way of air-to-ground visual cues. The intent was not to update these simulators with the most current hardware available but to use what hardware was in the simulators at the time, note any limitations which could be eliminated or improved by recent hardware developments and substantiate these comments with engineering data. It should also be pointed out that the LAMARS simulator was designed primarily for air-to-air combat mission evaluations and handling quality studies. This simulator is not a training simulator but an engineering design and evaluation tool obtained primarily for handling qualities evaluation and air-to-air combat studies.

II. DESCRIPTION OF THE MAJOR SIMULATOR SUBSYSTEMS

Visual Display System

The visual display system of the LAMARS consists of a 20-foot diameter spherical projection screen fixed to the cockpit. Within the sphere are two projectors which display the visual information to the pilot (Figure 1). One projector mounted above and to the rear of the cockpit projects the high resolution detailed terrain image while the other projector, located above and slightly to the rear of the target projector, referred to as the sky/earth projector, projects a passive blue sky, green earth horizon display throughout the entire sphere.

As was mentioned previously, the pilot is at the center of the spherical screen, which helps eliminate distortion of the projected images. As a result, the target projector and the sky-earth projector are off center and the projectors themselves have been specially designed to compensate for this effect. The target projector has a docus servo which moves the cathode-ray tube (CRT) in and out with respect to the projection lens. By knowing the

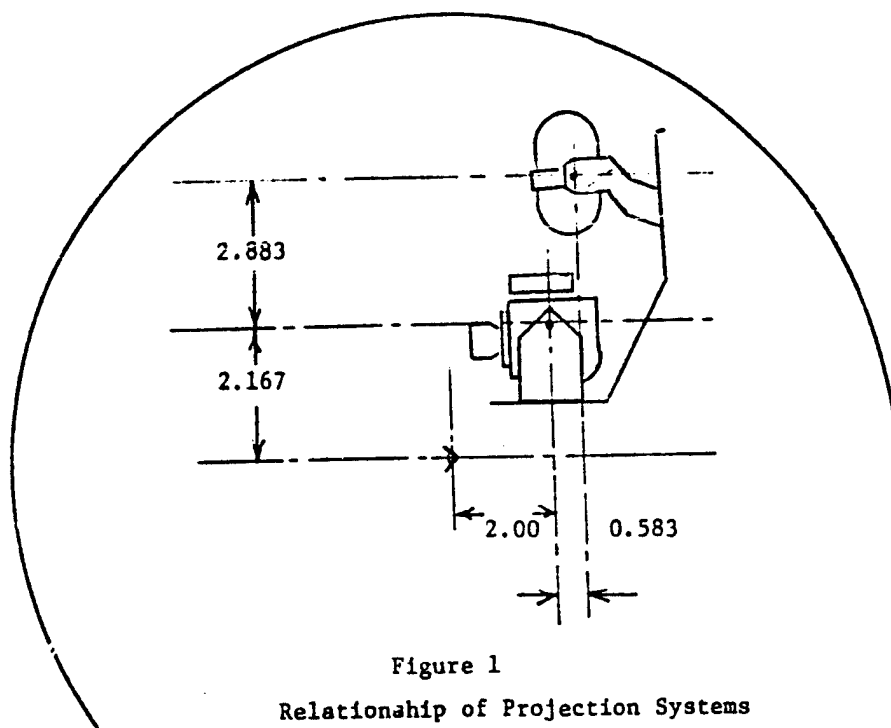


Figure 1
Relationship of Projection Systems

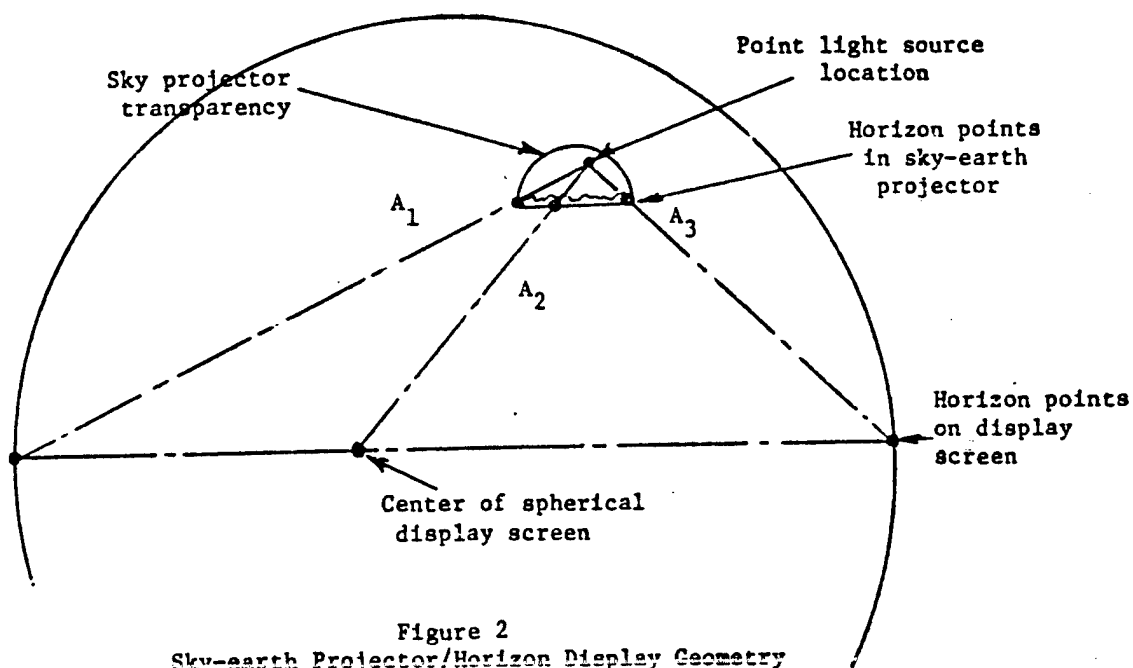


Figure 2
Sky-earth Projector/Horizon Display Geometry

position of the projector with respect to the center of the sphere and also the azimuth and elevation of the projector, the "throw distance" or focal length from the projector to the screen can be calculated and used to drive the focus servo. The target projector also has a "keystone" correction provision in its raster control circuitry to compensate for the varying angles at which the projected image impinges on the surface of the screen. The sky/earth projector has a mechanism inside the twin hemispherical transparent domes which can position the point light source of the "sky" lamp and "earth" lamp, to the proper x, y, and z coordinates within the transparency so that the projected horizon line is always correctly oriented on the screen without distortion. This is demonstrated in Figure 2.

Terrain Model/Computer Image Generation

The image which is presented to the pilot is generated by a large scale model of a typical terrain area. This three dimensional full color terrain model has its own gantry supported, optical probe equipped television camera positioned by precision servo drive systems commanded by computer signals. These computer signals are a result of the aircraft orientation and provide properly oriented visual cues to the pilot. Also, computer generated images (CGI) can be programmed and superimposed over the terrain image to provide moving targets, Forward Air Controller (FAC) aircraft, computer generated heads-up displays, etc; however, their utility using a television display system is questionable (see Appendix I). The resultant video picture is transmitted to the cathode-ray tube of the target projector and is projected onto the spherical screen. The sky/earth horizon display and the horizon in the video picture are synchronized by superposition of the terrain image and sky/earth image on the dome. This provides a 60° diagonal forward field-of-view displaying high resolution terrain imagery and also a wide field of view horizon display to produce the necessary peripheral cues needed to help the pilot in judging sink rate, roll position, roll rates, etc.

Projection Capability

The requirement for a 60° diagonal area-of-interest projection beam which can be positioned anywhere within the field-of-view of the pilot arose from the desire to have a display capability which could provide the visual requirements for such tasks as air-to-ground weapon delivery, terrain following/avoidance, and take off/landing work. The resulting display capability represents a compromise between a full field of view terrain image and what was believed to be a realistic achievement based on a thorough review of existing hardware.

The requirement of a large field of view results from the new aircraft designs, which to properly simulate, require an unlimited view in both azimuth and elevation. The LAMARS simulator was able to achieve a 266° field of view in azimuth and a 108° field of view in elevation. These values were not quite as large as was desired (280° in azimuth, 150° in elevation) but represent the largest field of view possible, limited only by the cockpit

gimbal system and the projector system mounting configuration.

III. THE AIR TO GROUND WEAPONS DELIVERY TASK

In performing an air-to-ground task in the LAMARS there are three methods in which to display the high resolution terrain image. The first, and most basic is to fix the 60° area of interest to the x-axis of the aircraft. In doing this, the rectangular AOI will remain fixed in the front window of the cockpit. The picture within the field of view will change solely as a result of the aircraft orientation. This provides a forward fixed scene but does not permit the pilot to perform his typical roll maneuver when attacking a ground target since the target at that point will be out of the area of interest. Secondly, it does not utilize the large field-of-view capability of the spherical projection screen.

The second approach would be to fix the visual scene to a specific target. This would permit the 60° diagonal terrain image to appear anywhere in the dome based on the relative positions of the aircraft and the target. For example, if the pilot was flying into the target the terrain image would be directly in front of the pilot. If the pilot flew by the target the terrain image would move off to the side and on back to the rear of the dome and finally disappear from view. This is certainly more realistic than the first case but does not permit the pilot to do terrain following and target search since he knows that if he keeps the terrain image directly in front of him he will eventually fly directly over the target. To help eliminate the problems associated with the above methods, an area-of-interest, head-slaved visual display system was developed.

Head Slaved Visual System

To develop a head-slaved visual system, one must first be able to monitor the pilot's head position. This should be done in a manner which is completely unobtrusive to the pilot. Once the head position is known and calibrated with respect to a fixed reference point in the cockpit then drive logic and its associated software can be developed.

The hardware used in this experiment was a state-of-the-art helmet sight system. Sensor surveying units were rigidly fixed to the cockpit and aimed in the direction of the pilot's head. These units emit fan-like beams of infrared light rotating at a constant velocity. The infrared beams sweep over a reference photo sensor and two pairs of helmet mounted photo sensors, one pair on each side of the helmet. The time intervals between the pulses from the helmet photo sensors and the reference pulse are a measure of the pilot's head orientation relative to the body axis of the aircraft. The outputs of the photo sensors are transmitted to the helmet sight computer where the angular computations are performed and converted into azimuth and elevation information. The accuracy of the system is one-half degree CEP.

As was shown before, the target projector and the pilot's head are at

different locations within the sphere. As a result, to point the center of the image projected by the target projector to intercept the point directly in front of the pilot's head, a geometrical transformation is required.

At the same time, it requires that the distance from the target projector lens to the sphere be calculated to drive the focus servo. Because the target projector probe is not as responsive as the angular position servos of the visual probe, the angles of the head are then recalculated using a similar geometric transformation and the position feedback signals from the target projector. The calculated head position and the aircraft position are combined to form a resultant drive signal to the visual probe servos. The probe servos are lead compensated to provide better response. This arrangement permits the pilot to look about within the sphere and view a correctly oriented visual segment.

Other software has also been added to help make the visual image seen by the pilot more realistic. One of these is a dead zone filter which is used to help eliminate small amplitude, high frequency head movements from placing jitter into the system. The dead zone requires that the pilot move his head some predetermined delta value before the target projector repositions itself. This allows the pilot to slowly scan the terrain image without constantly repositioning the projector. It also acts to keep out high frequency noise which could result from calculation errors in the digital machine.

IV. EVALUATION TEST PLAN

The LAMARS air-to-ground visual system was evaluated by six pilots from the Tactical Air Command, Aeronautical Systems Division and the Navy. These pilots were currently rated in various high performance aircraft and were highly experienced in the air-to-ground task.

Each pilot flew the device for a total of eight one-hour sorties. The pilots performed the following weapon deliveries:

1. 10° skip bomb
2. 15° low angle bomb
3. 15° low angle strafe
4. 20° low angle dive bomb
5. 30° high angle strafe
6. 30° dive bomb
7. 45° dive bomb

At weapon release the following data was recorded:

1. aircraft position, velocity, angles, angular velocities

2. weapon miss distance (bomb) or mean impact point (gun)

A number of tactical targets were used during the sorties.

1. range target
2. gravel pit
3. suspension bridge
4. dam pumphouse
5. railroad overpass
6. small house
7. runway lights
8. highway intersection
9. oil tank
10. SAM site
11. parked aircraft
12. bulldozer
13. airport terminal

The aircraft could deliver either a MK82 bomb or deliver gunfire with a 30mm gun.

A pilot questionnaire was used to gather information on:

1. opinion of weapon delivery performance
2. ease of target acquisition
3. comments on AOI display
4. comments on simulation realism
5. other pertinent comments

The following numerical data was computed for each pilot/target/delivery conditions:

1. mean along track miss distance
2. standard deviation, along track miss distance
3. mean cross track miss distance
4. standard deviation, cross track miss distance
5. correlation coefficient, along track and cross track miss distance
6. median radial error
7. maximum miss distance

8. minimum miss distance
9. the number of hits long, short, left, and right of target.

Engineering tests were also conducted on the LAMARS visual system itself. These tests included

1. System Static Resolution/Modulation Transfer Function
2. Image Generator Static Resolution/Modulation Transfer Function
3. Display Static Resolution/Modulation Transfer Function
4. Image Generator Dynamic Resolution/Modulation Transfer Function
5. Display Brightness
6. Display Grayscale
7. Display Contrast
8. Display Shading
9. System Brightness
10. System Grayscale
11. System Contrast
12. System Shading
13. System Geometric Distortion
14. AOI Field-of-View
15. AOI Dynamic Envelope Size
16. AOI Edge Transition Quality
17. Target Image Location Dynamic Lag
18. System Rate Accuracy
19. System Design Data

V. PILOT EVALUATION RESULTS

The pilots were debriefed at the end of each 60 minute sortie and were asked to evaluate the task with respect to the LAMARS head-slaved visual system. The following comments and their cause were noted as being common comments of all pilots.

Attributes

1. Diversified Scene Content - The model board type of image generation is unique in its ability to produce high quality scene detail. The detail has good texture and is significantly more detailed than computer generated imagery.

2. Smooth Projection Surface - The spherical dome does not have discrete display surfaces as do CRT type displays. Because of this the scene can transition throughout the $266^{\circ} \times 108^{\circ}$ viewing area with smooth continuity.

3. Head Slaved AOI is Promising - The pilots all felt that the head slaved area-of-interest display was the best of the three projection methods within the LAMARS but felt it was presently limited due to other hardware shortcomings presently found in the rigid model visual system.

Limitations

1. AOI Too Small - The rigid model visual system presently is limited to a 60° diagonal area of interest (48° wide by 36° high). This is inadequate for many tactical maneuvers. For example, to attack a ground target or to turn from base leg to final during landing requires a minimum of 90° field of view. A navigational land mark is used to guide the aircraft and when the target or runway appears over the left wing a turn or roll maneuver is executed. With a 48° wide area of interest, the target and navigational aid are not within the same visual segment. In the head slave mode, the pilot can command the probe/target projector to pick up the visual cues he needs to perform the task. The pilot, however, is now placed in a situation where he is performing unrealistic head movements which degrades his performance. A minimum of 120° diagonal (96° wide by 72° high) area of interest is required to perform a realistic tactical maneuver using a head slaved visual system.

2. 47° Pitch Restriction Too Small - The AFFDL rigid model visual system has a 47° down pitch hardware limit on the visual probe. When this limit is exceeded the visual segment is obscured by a pseudo-cloud cover. In high angle tactical maneuvers, since the head and aircraft now command the probes pitch position, these limits can be easily exceeded. It is quite distracting and unnatural to lose the target near weapon release. An unlimited pitch axis on the probe is an "off-the-shelf" modification which can be easily performed on the AFFDL system.

3. Uncertain Scoring Accuracy - Scoring accuracy using head-slaving is presently uncertain because it was not possible to calibrate it prior to pilot evaluation. Weapon delivery data, at first look, appears to indicate that there is a dynamic lag using head slaving which could cause impact errors on the order of several hundred feet given the approach speeds and dive angles used in this simulation.

4. Horizon Mismatch - The sky/earth horizon display did not always match the horizon display from the rigid model terrain board image. The reason for this is that the two were driven from separate loops, one was programmed for flat earth (terrain board image) while the other was for spherical earth (sky/earth display) and the sky-plate used by the probe to prevent viewing the probe/gantry in the mirrors can displace the apparent horizon location at high altitude or near the mirrors surrounding the board. Further programming improvements will alleviate this problem.

The pilots in general found the combination of the small field-of-view and the pitch restriction on the probe produced unrealistic piloting conditions. They did feel that if these two problems were alleviated/eliminated, a viable air-to-ground display could result.

VI. ACKNOWLEDGEMENTS

Obviously, to conduct a program of this complexity, many people must be involved. The general concepts described in this paper and the software equations were developed by the authors. These equations, however, were in Fortran IV and had to be rewritten in machine language for implementation on the LAMAR simulator. Mr. William J. Waldron, who is responsible for the operation and programming of the LAMARS, spent considerable time with much personal involvement in getting the head slaved visual system operational. Appreciation must be given to Mr. Waldron for his effort in this project and also to Mr. James A. Eicher, who was responsible for tying in the visual system.

REFERENCES

1. Haas, R.L., H.E. Hotz, and G.R. Mills. "The Large Amplitude Multi-Mode Aerospace Research (LAMAR) Simulator". AIAA Paper No. 73-922, AIAA Visual and Motion Simulator Conference, Palo Alto, California. Sept 10-12, 1973.
2. Gressang, R.V., D. Kugel, W. Waldron, J. Eicher. "Drive Signals for Positioning an Area of Interest Visual Display". AFFDL TM 63-13-FGD, February 1976.
3. Kugel, Daniel L. and R.V. Gressang. "Air-to-Ground Visual Display System". Proceedings, AIAA Visual and Motion Simulation Conference, Dayton, Ohio, 26-28 April 1976.

APPENDIX I

TELEVISION DISPLAY RESOLUTION

When a television display is used, one must keep in mind that each visual picture is obtained from a series of raster lines painted by a cathode ray gun. The more lines a display is able to paint across the face of the screen, the more accurate will be the resulting picture. Using a typical 60° field of view probe and a television display whose resolution is 700 effective lines, one can determine at what distance a target of known dimensions will be able to be recognized.

If N = number of effective resolved lines across the sensor's FOV

M = number of lines across the target for identification/recognition

L = maximum dimension of target

S = slant range

then using

$$\text{probe FOV required} = \frac{2N}{M} \tan^{-1} \frac{L}{2S} \quad (\text{I-1})$$

if 8 lines are required for recognition of a moving target whose maximum dimension is 30 feet, the effective resolution of the display is 700 lines, and the probe FOV is 60° , then

$$S = \frac{L}{2 \tan \left\{ \text{probe FOV} \cdot \frac{M}{2N} \right\}} \quad (\text{I-2})$$

$$S \approx 2500 \text{ ft} \sim 1/2 \text{ mile.}$$

It is obvious that in unlimited visibility, a 30 ft aircraft can be seen for several miles. Using the same information as above, a 30 ft airplane at two miles would be painted by less than two scan lines resulting in only a single dot on the screen.

APPENDIX II

SOFTWARE EQUATIONS

The block diagram for the head slaved visual system is shown in Figure 3. In this appendix, the important software equations are described.

A. Head to Target Projector Transformation

Azimuth

$$\psi^* = 180.0 - \psi_H$$

$$R_\psi^2 = [10.0]^2 + [2.0]^2 - 2(10.0)(2.0) \cos \psi^*$$

Therefore

$$R_\psi = [104.0 - 40.0 \cos (180.0 - \psi_H)]^{1/2} \quad (\text{II-A-1})$$

LAMARS HEAD-SLAVED VISUAL SYSTEM

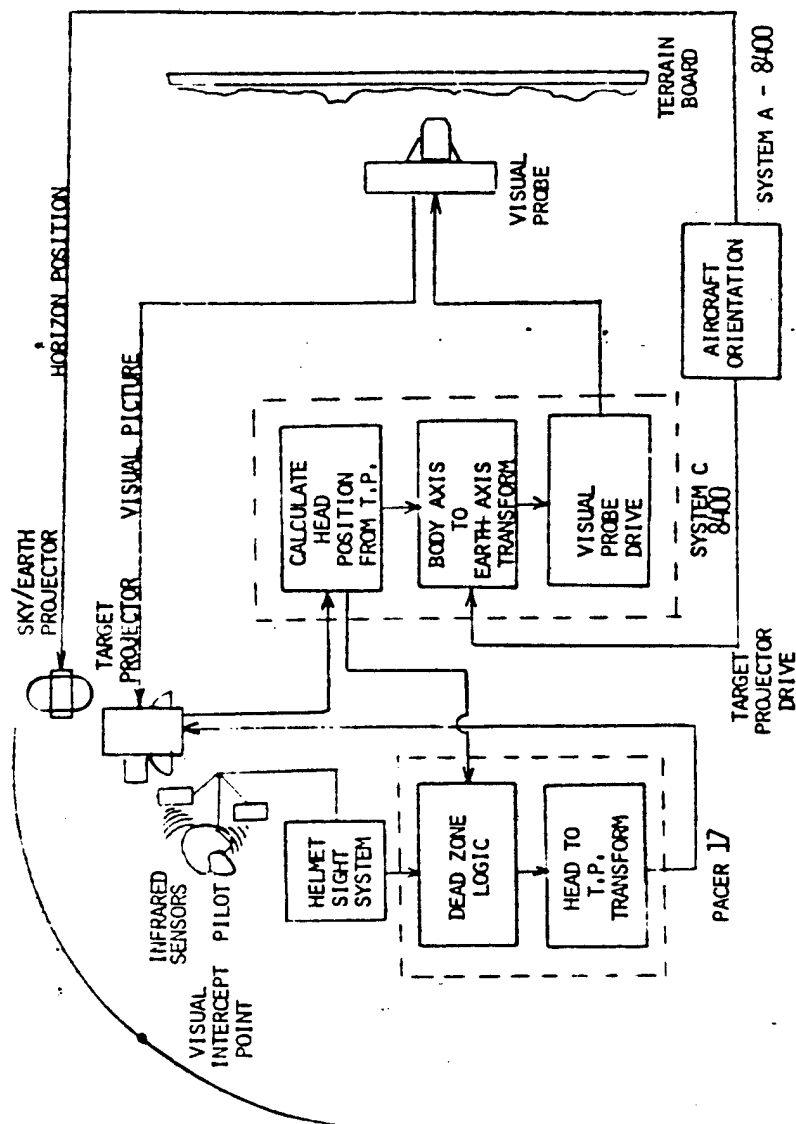


Figure 3

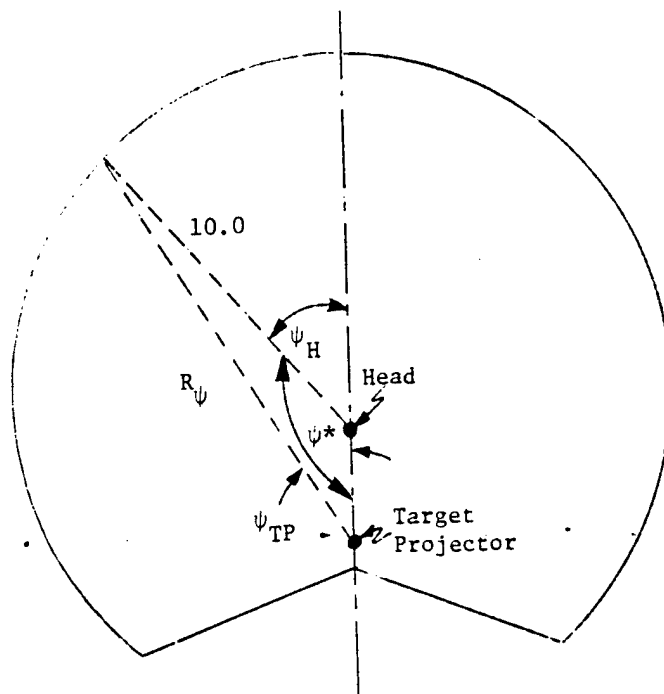


Figure 4

Also,

$$\psi_{TP} = \cos^{-1} \left[\frac{(2.0)^2 + R_\psi^2 - (10.0)^2}{2(2.0)R_\psi} \right]$$

which reduces to

$$\psi_{TP} = \cos^{-1} \left[\frac{R_\psi^2 - 96.0}{4.0R_\psi} \right] \quad (\text{II-A-2})$$

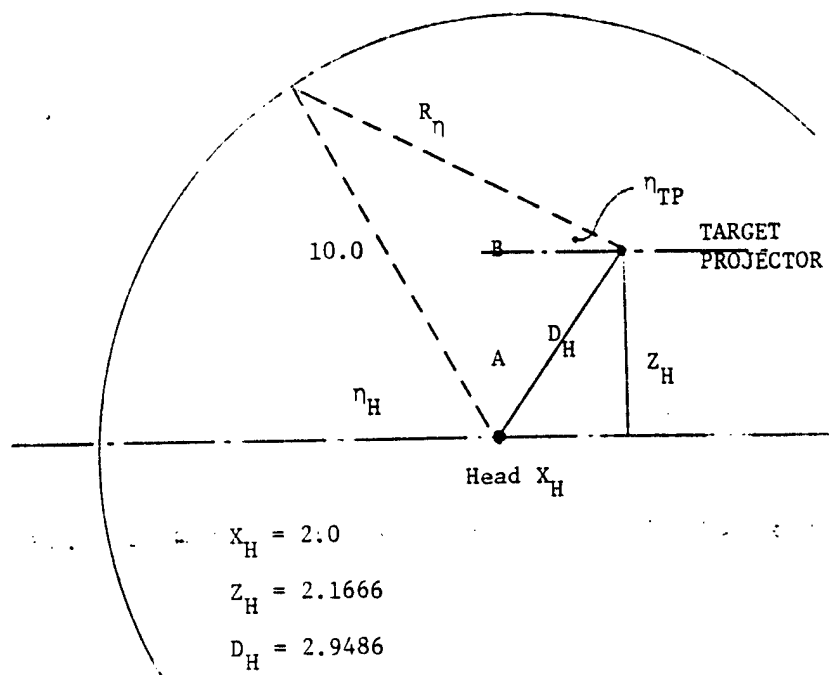


Figure 5

Then

$$\eta_H + A + 47.29^\circ = 180^\circ$$

or

$$A = 132.71 - \eta_H$$

$$R_\eta^2 = [10.0]^2 + [2.9486]^2 - 2[10.0][2.9486] \cos A$$

$$R_\eta = [108.6944 - 58.9727 \cos (132.71 - \eta_H)]^{1/2} \quad (\text{II-A-3})$$

Also

$$-\eta_{TP} + B + 42.71 = 90^\circ$$

or

$$\eta_{TP} = B - 47.29^\circ$$

$$B = \cos^{-1} \left[\frac{R_{\eta}^2 + (2.9486)^2 - (10.0)^2}{2(2.9486)R_{\eta}} \right]$$

which reduces to

$$B = \cos^{-1} \left[\frac{R_{\eta}^2 - 91.3055}{5.8973R_{\eta}} \right]$$

and

$$\eta_{TP} = \cos^{-1} \left[\frac{R_{\eta}^2 - 91.3055}{5.8973R_{\eta}} \right] - 47.29^{\circ} \quad (\text{II-A-4})$$

Using the body axis coordinates, x, y, z, then the location of the head with respect to the target projector is

$$x_H = 2.0$$

$$y_H = 0.0$$

$$z_H = -2.1666$$

The components of the vector from the head to the screen can be given by

$$x_p = R \cos \psi_H \cos \eta_H \quad (\text{II-A-5})$$

$$y_p = R \sin \psi_H \cos \eta_H \quad (\text{II-A-6})$$

$$z_p = R \sin \eta_H \quad (\text{II-A-7})$$

The throw distance to the screen is then

$$R_{\hat{f}} = [(x_H + x_p)^2 + (y_H + y_p)^2 + (z_H + z_p)^2]^{1/2} \quad (\text{II-A-8})$$

B. Calculated Head Angles from Target Projector Feedbacks

To determine the calculated head angles based on the position feedback signals from the target projector, ψ_{TP} and η_{TP} are replaced with $\psi_{TP_{FB}}$

and $\eta_{TP_{FB}}$ respectively, in equations II-A-1 through II-A-4.

The results are

$$R_{\psi_{CAL}} = 2.0 \cos \psi_{TP_{FB}} + 0.5[392.0 + 8.0 \cos \psi_{TP_{FB}}]^{1/2} \quad (II-B-1)$$

$$\psi_{H_{CAL}} = \cos^{-1} \left[\frac{R_{\psi_{CAL}}^2 - 104.0}{40.0} \right] \quad (II-B-2)$$

$$R_{\eta_{CAL}} = 2.9486 \cos (\eta_{TP_{FB}} + 47.29^\circ) + 0.5\{17.39[1 + \cos(2\eta_{TP_{FB}} + 94.58^\circ)] + 365.22\}^{1/2} \quad (II-B-3)$$

$$\eta_{H_{CAL}} = 132.71 - \cos^{-1} \left[\frac{108.6944 - R_{\eta_{CAL}}^2}{58.9727} \right] \quad (II-B-4)$$

C. Axis Transformations

Given a set of earth fixed axes, tangent to the surface of the earth, z axis down, and the aircraft Euler angles (body axis, ϕ , θ , ψ) and the head angles with respect to the body axis (η_H , ψ_H , $\phi_H = 0$), the various axis transformation matrices are desired. The transformation matrix, C_E^B , from earth axis to body axis is given in terms of the aircraft Euler angles, ϕ , θ , ψ by

$$C_E^B = \begin{bmatrix} c\phi c\psi & c\theta s\psi & -s\theta \\ s\phi s\theta c\psi - c\phi s\psi & s\phi s\theta s\psi + c\phi c\psi & s\phi c\theta \\ c\phi s\theta c\psi + s\phi s\psi & c\phi s\theta s\psi - s\phi c\psi & c\phi c\theta \end{bmatrix} \quad (II-C-1)$$

where c denotes cosine and s denotes sine. The transformation matrix, C_B^H , from body axis to head axis, is given by

$$C_B^H = \begin{bmatrix} c\eta_H c\psi_H & c\eta_H s\psi_H & -s\eta_H \\ -s\psi_H & c\psi_H & 0 \\ s\eta_H c\psi_H & s\eta_H s\psi_H & c\eta_H \end{bmatrix} \quad (II-C-2)$$

where ψ_H is the head azimuth and η_H is the head elevation (Euler angles) relating head to body axes. For our purposes ϕ_H (head roll) is not measured and is identically equal to zero.

The transformation from earth to head is then calculated using

$$C_E^H = C_B^H C_E^B \quad (II-C-3)$$

$$= \begin{bmatrix} c\theta_{EQ} c\psi_{EQ} & c\theta_{EQ} s\psi_{EQ} & -s\theta_{EQ} \\ (s\phi_{EQ} s\theta_{EQ} c\psi_{EQ} & (s\phi_{EQ} s\theta_{EQ} s\psi_{EQ} & s\phi_{EQ} c\theta_{EQ} \\ -c\phi_{EQ} s\psi_{EQ}) & +c\phi_{EQ} c\psi_{EQ}) \\ (c\phi_{EQ} s\theta_{EQ} c\psi_{EQ} & (c\phi_{EQ} s\theta_{EQ} s\psi_{EQ} & c\phi_{EQ} c\theta_{EQ} \\ + s\phi_{EQ} s\psi_{EQ}) & -s\phi_{EQ} c\psi_{EQ}) \end{bmatrix} \quad (II-C-4)$$

where

$$\theta_{EQ} = \sin^{-1} [c\eta_H c\psi_H s\theta - c\eta_H s\psi_H s\phi c\theta + s\eta_H c\phi c\theta] \quad (II-C-5)$$

$$\phi_{EQ} = \tan^{-1} \left[\frac{s\psi_H s\theta + c\psi_H s\phi c\theta}{-2\eta_H c\psi_H s\theta + s\eta_H s\psi_H s\phi s\theta + c\eta_H c\psi_H c\theta} \right] \quad (II-C-6)$$

$$\psi_{EQ} = \tan^{-1} \left[\frac{(c\eta_H c\psi_H c\theta c\psi + c\eta_H s\psi_H s\phi s\theta s\psi) + (c\eta_H s\psi_H c\phi c\psi - s\eta_H c\phi s\theta s\psi + s\eta_H s\phi s\psi)}{(c\eta_H c\psi_H c\theta c\psi + c\eta_H s\psi_H s\phi s\theta c\psi) + (-c\eta_H s\psi_H c\phi s\psi - s\eta_H c\phi s\theta c\psi - s\eta_H s\phi s\psi)} \right] \quad (II-C-7)$$

D. Dead Zone Logic

This is a nonlinear filter realized by two logic tests which allows the target projector to be moved only by head position. Random small amplitude noise is filtered out permitting a smooth projector response. The position and rates of the head can be calculated using the following equations.

$$\text{Elevation position limit} = \text{ABS}(\eta_H(\text{new}) - \eta_H(\text{old})) \quad (\text{II-D-1})$$

$$\text{Azimuth position limit} = \text{ABS}(\psi_H(\text{new}) - \psi_H(\text{old})) \quad (\text{II-D-2})$$

$$\text{Elevation rate limit} = \text{ABS}(\eta_H(\text{new}) - \eta_{H_{\text{CAL}}}(\text{old})) \quad (\text{II-D-3})$$

$$\text{Azimuth position limit} = \text{ABS}(\psi_H(\text{new}) - \psi_{H_{\text{CAL}}}(\text{old})) \quad (\text{II-D-4})$$

Using a 25 sample/second update rate, the required position and rate limits can be calculated. As long as the pilot does not exceed either the position or rate limits, then the target projector will remain fixed. As soon as either the rate limit or the position limit is exceeded, the target projector will move to a new position based on the commands from the head sensing unit. This allows for small head motions to be made and not affect the target projector and also helps eliminate noise on the signals from driving the target projector.

E. Visual Probe Drive Equations

The visual probe is also required to be slaved to the pilot's head (line of sight). The visual probe orientation is determined by the transformation matrix from earth axis to head axis, where

C_E^H , θ_{EQ} , ϕ_{EQ} and ψ_{EQ} are calculated as in section C.

The probe compensation required was based on the target projector transfer function of

$$H(s) = \frac{K}{s^2 + 2(.7)(25)s + (25)^2}$$

The probe was compensated to have a transfer function of

$$H(s) = \frac{K(s+55)}{s^2 + 2(.7)(25)s + (25)^2}$$

with the compensated probe having a 6% overshoot and a 6 ms lag. The compensation was performed in the digital computer, using the equation

$$\text{DRIVE} = \frac{(\text{ANGLE}_i - \text{ANGLE}_{i-1})}{\Delta T} * \frac{1}{55} + \text{ANGLE}_i.$$

The angles were nondimensionalized and ΔT was about 16 ms.

IN-FLIGHT SIMULATION STUDY OF DECOUPLED
LONGITUDINAL CONTROLS FOR THE APPROACH
AND LANDING OF A STOL AIRPLANE¹

By Edward Seckel and Benjamin Feinreich

Princeton University

SUMMARY

A simulation study of a powered lift STOL transport having decoupled longitudinal controls for the approach and landing flight phases has been conducted on the Princeton Navion in-flight simulator. In the decoupled control concept, the natural interacting airplane responses (combined pitch attitude, speed, and flight path angle changes for fore and aft stick motion, for example) are suppressed, and the pilot operates a separate control lever for each variable.

In this study, fore and aft control column motion produced changes in flight path angle without changing attitude or speed; the throttle commanded speed changes independent of attitude or flight path angle; and the pitch trim wheel allowed independent pitch attitude changes.

Landings were made out of various typical STOL straight and segmented approaches using ILS and precision optical guidance. The flying qualities were judged to be very favorable, although a short period of adjustment to the unconventional constant-attitude, constant-speed flare was required. The precise control over flight path resulted in small touchdown point dispersion along with consistently low sink rates.

SYMBOLS

C_L lift coefficient

C_{L_∞} lift coefficient out of ground effect

$d\gamma/du$ change of flight path with speed, thrust constant, deg/kt

¹ This study was performed for NASA, Langley Research Center, under Contract NAS1-13502.

g	acceleration due to gravity, m/sec^2 , ft/sec^2
h	altitude, m, ft
\dot{h}	vertical velocity, m/sec, ft/sec
M_u	pitch acceleration derivative due to speed, $\frac{1}{I_y} \frac{\partial M}{\partial u}$, rad/sec^2 per m/sec, rad/sec^2 per ft/sec
M_{u_c}	pitch acceleration derivative due to forward speed command input, $\frac{1}{I_y} \frac{\partial M}{\partial u_c}$, rad/sec^2 per m/sec, rad/sec^2 per ft/sec
M_α	pitch acceleration derivative due to angle of attack, $\frac{1}{I_y} \frac{\partial M}{\partial \alpha}$, rad/sec^2 per rad
$M_{\dot{\alpha}}$	pitch acceleration derivative due to rate of change of angle of attack, $\frac{1}{I_y} \frac{\partial M}{\partial \dot{\alpha}}$, rad/sec^2 per rad/sec
M_γ	pitch acceleration derivative due to flight path angle, (used when the equations of motion are written in u , γ , θ , rather than the more common u , α , θ ; for unaugmented airplane $M_\gamma = -M_\alpha$), $\frac{1}{I_y} \frac{\partial M}{\partial \gamma}$, rad/sec^2 per rad
$M_{\dot{\gamma}}$	pitch acceleration derivative due to rate of change of flight path angle, $\frac{1}{I_y} \frac{\partial M}{\partial \dot{\gamma}}$, rad/sec^2 per rad/sec
M_{γ_c}	pitch acceleration derivative due to flight path command input, $\frac{1}{I_y} \frac{\partial M}{\partial \gamma_c}$, rad/sec^2 per rad
M_θ	pitch acceleration derivative due to pitch attitude, $\frac{1}{I_y} \frac{\partial M}{\partial \theta}$, rad/sec^2 per rad
M_{θ_c}	pitch acceleration derivative due to pitch command input, $\frac{1}{I_y} \frac{\partial M}{\partial \theta_c}$, rad/sec^2 per rad/sec

$M_{\dot{\theta}}$	pitch rate damping, $\frac{1}{I_y} \frac{\partial M}{\partial \dot{\theta}}$, rad/sec ² per rad/sec
p_g	roll rate equivalent of linearly spanwise distributed vertical gust velocity component, rad/sec, deg/sec
q	pitch rate, rad/sec, deg/sec
s	Laplace transform variable
S	wing reference area, m ² , ft ²
T	thrust, N, lb
u	airspeed perturbation, m/sec, ft/sec
u_g	fore and aft gust velocity component, m/sec, ft/sec
v_g	side gust component, m/sec, ft/sec
V_o	trim airspeed, knots, m/sec, ft/sec
w	vertical speed perturbation, m/sec, ft/sec
w_g	vertical gust velocity component, m/sec, ft/sec
X_u	longitudinal acceleration derivative due to forward speed, $\frac{1}{m} \frac{\partial X}{\partial u}$, 1/sec
X_{u_c}	longitudinal acceleration derivative due to forward speed command, $\frac{1}{m} \frac{\partial X}{\partial u_c}$, 1/sec
X_{α}	longitudinal acceleration derivative due to angle of attack, $\frac{1}{m} \frac{\partial X}{\partial \alpha}$, m/sec ² per rad, ft/sec ² per rad
X_{γ}	longitudinal acceleration derivative due to flight path angle, $\frac{1}{m} \frac{\partial X}{\partial \gamma}$, m/sec ² per rad, ft/sec ² per rad

X_{γ_c}	longitudinal acceleration derivative due to flight path command input, $\frac{1}{m} \frac{\partial X}{\partial \gamma_c}$, m/sec ² per rad, ft/sec ² per rad
X_{θ}	longitudinal acceleration derivative due to pitch attitude $\frac{1}{m} \frac{\partial X}{\partial \theta}$, m/sec ² per rad, ft/sec ² per rad
X_{θ_c}	longitudinal acceleration derivative due to pitch attitude command input, $\frac{1}{m} \frac{\partial X}{\partial \theta_c}$, m/sec ² per rad, ft/sec ² per rad
Z_u	vertical acceleration derivative due to forward speed, $\frac{1}{m} \frac{\partial Z}{\partial u}$, 1/sec
Z_{u_c}	vertical acceleration derivative due to forward speed command input, $\frac{1}{m} \frac{\partial Z}{\partial u_c}$, 1/sec
Z_w	vertical acceleration derivative due to vertical speed, $\frac{1}{m} \frac{\partial Z}{\partial w}$ ($\equiv Z_{\alpha}/V_o$), 1/sec
Z_{α}	vertical acceleration derivative due to angle of attack, $\frac{1}{m} \frac{\partial Z}{\partial \alpha}$, m/sec ² per rad, ft/sec ² per rad
Z_{γ}	vertical acceleration derivative due to flight path angle, $\frac{1}{m} \frac{\partial Z}{\partial \gamma}$, m/sec ² per rad, ft/sec ² per rad
Z_{γ_c}	vertical acceleration derivative due to flight path angle command input, $\frac{1}{m} \frac{\partial Z}{\partial \gamma_c}$, m/sec ² per rad, ft/sec ² per rad
Z_{θ}	vertical acceleration derivative due to pitch attitude, $\frac{1}{m} \frac{\partial Z}{\partial \theta}$, m/sec ² per rad, ft/sec ² per rad
Z_{θ_c}	vertical acceleration derivative due to pitch command input, $\frac{1}{m} \frac{\partial Z}{\partial \theta_c}$, m/sec ² per rad, ft/sec ² per rad
α	angle of attack, rad, deg

γ	flight path angle, rad, deg
γ_c	flight path angle command, rad, deg
$\Delta()$	perturbation from trim condition
δ_c	fore and aft control column deflection, cm, in.
δ_f	flap deflection, rad, deg
δ_{sp}	spoiler deflection, cm, in.
δ_t	horizontal tail deflection, rad, deg
δ_{th}	throttle deflection, cm, in.
θ	pitch attitude, rad, deg
$\dot{\theta}$	pitch rate, $\frac{d\theta}{dt}$, rad/sec, deg/sec
θ_c	pitch command, rad, deg

INTRODUCTION

Piloting Problems of STOL Airplanes

The approach and landing is probably the most demanding task performed by the pilot of a transport airplane. It is even more so with an STOL aircraft that is landed out of a steeper glide slope with a higher accuracy requirement on touchdown dispersion. The problem is further complicated in the case of powered lift machines by their poor handling qualities in approach and landing. A stability augmentation system (SAS) is usually necessary to provide acceptable flying qualities. However, some deficiencies, mainly in the longitudinal plane, exist even with the stability augmentation. The most important problem is a poor flight path angle response to attitude changes which makes it difficult to use pitch attitude for precision control of the flight path. In the powered lift airplane a sizable flight path angle response is associated with the throttle, but if jet engines are the power plants, there is a time lag between the throttle movement and the change in engine thrust. This time lag creates problems in attempting to use the throttle as a precision flight path control. The inadequacy of either the control column or the throttle alone in the control of flight path angle necessitates the use of both. Typically all three flight variables,

flight path angle, forward speed, and pitch attitude change in response to control column displacement; and flight path and speed respond to throttle movement. Consequently, the pilot has to coordinate action on both levers in order to control the three coupled variables. All this results in the deficient flying qualities. Adverse ground effect and turbulence cause a substantial increase in pilot workload.

The above observations are substantiated by Reference 1 in which a fixed base simulator study that has been conducted by NASA Langley Research Center (LRC) is described. The purpose of this study was to determine the flight characteristics during the approach and landing of a representative STOL transport airplane having a high wing and equipped with an external-flow jet flap in combination with four high-bypass-ratio fan-jet engines (see Figure 1). Conventional stability augmentation systems (SAS) have been applied to obtain satisfactory handling qualities.

The Concept of Decoupling

Decoupled longitudinal controls have been suggested in Reference 2 in an attempt to improve the handling qualities of the airplane treated in Reference 1. The essence of this concept is to make each one of the three flight variables respond only to one cockpit control. In this study, the column was chosen to affect flight path angle and not to change speed or attitude. The throttle handle was chosen to control speed without affecting flight path or attitude, and a pitch-trim thumbwheel controlled attitude without changing flight path angle or speed. Provided that the quality and authority of each of the flight variable responses to their appropriate cockpit levers are adequate, the pilot can use one lever to control one flight variable.

In landing such a decoupled STOL airplane, speed and pitch attitude are stabilized early in the approach and the pilot does not have to concern himself with their active control any more. He can concentrate on controlling flight path angle by using the column only. With the excellent flight path response to column that was provided, this resulted in a very significant improvement of flying qualities which is especially pronounced in the presence of adverse ground effect and turbulence.

The price that has to be paid for this improvement is an unconventional and complex flight control system that in the configuration suggested in Reference 2, employed four feedback variables (u , α , θ , q) and four active control elements: throttle, symmetric spoiler, flap, and horizontal tail.

Ground Simulator Study of Decoupled Controls

A fixed base, ground simulator study of decoupled controls for the externally blown flap (EBF) STOL in the approach and landing was the subject of Reference 2. An improvement in flying qualities of the longitudinally decoupled airplane in comparison to the conventionally augmented one was reported. Visual cues in the simulator were obtained by a closed circuit television system. However, the use of this system for image generation caused difficulties in sensing altitude and altitude rate resulting in higher sink rates at touchdown and higher touchdown point dispersion along the runway than would be expected in real flight.

The same control concept has been tested on a moving base simulator (Reference 3) which increased the realism of the simulation by providing motion cues. However, a closed circuit television generated image of the same type that had been used on the fixed base simulator, was also used on the moving base simulator. Another deficiency of the moving base simulator resulted from the wash-out circuits needed to prevent position, velocity, and acceleration limiting.

In-flight Simulation Study of Decoupled Controls

This paper summarizes an in-flight simulation study of the longitudinal decoupled controls for the same airplane that was the subject of References 1, 2, and 3. This study which was conducted on Princeton's in-flight simulator, was motivated by the positive results obtained in the ground simulators. The more realistic environment of the in-flight simulator with its real-world visual cues and its full unlimited motion (for the STOL approach and landing problem) produces an investigation tool that is superior to ground simulators.

Decoupled lateral controls were also introduced and tested in the ground simulators; however, the advantages offered over conventional SAS were not as significant as in the case of longitudinal decoupling; therefore, the in-flight investigation included only the latter.

A more detailed account of the in-flight study may be found in Reference 4.

DESCRIPTION OF THE EXPERIMENT

The In-flight Simulator

The in-flight simulator is shown in Figure 2. It is a "fly-by-wire" airplane with adjustable stability and control characteristics. In this test program they were adjusted to match the characteristics of an EBF STOL transport in an approach and landing configuration having decoupled longitudinal controls in some experiments and conventional SAS in others.

Tested Configurations

Five control configurations were tested in this program. Three of them were variations of decoupled controls:

- Steady State Decoupled (SSD)
- Completely Decoupled (CD)
- "Recoupled" (REC)

Two configurations were variations of conventional augmentation:

- Stability Augmentation System (SAS)
- Improved Stability Augmentation System (ISAS)

Following are the descriptions of those configurations and their characteristics.

Steady State Decoupled (SSD) Configuration - The decoupled longitudinal control scheme, described in Reference 2, was the main subject of this study. In this design, steady state, rather than complete decoupling was implemented. The term "steady state decoupling" means that whereas only one of the three flight variables (airspeed, flight path angle, or pitch attitude) exhibits a steady state change in response to the appropriate cockpit control lever, the other two variables may undergo transient variations.

The Navion in-flight simulator control assignments for this configuration were: column to control flight path angle with a beep trimmer on the left horn to trim γ ; throttle to control forward speed; pitch thumbwheel to control pitch attitude. The controls were mechanized such that the changes in the flight variables were proportional to control displacements from their trim positions. A constant $\pm 2^\circ/\text{sec } \gamma$ rate was associated with the beep trimmer.

The schematic structure of the decoupling control system that was employed in Reference 2 is shown in Figure 3.

This system may be represented by the following equations:

$$\begin{bmatrix} s - X_u & -X_\gamma & -X_\theta \\ Z_u/V_o & (s + Z_\gamma/V_o) & Z_\theta/V_o \\ -M_u & -(M_\dot{\gamma} s + M_\gamma) & (s^2 - M_\theta s - M_\theta) \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta \gamma \\ \Delta \theta \end{bmatrix} = \begin{bmatrix} X_{uc} & X_{\gamma c} \\ -Z_{uc}/V_o & -Z_{\gamma c}/V_o & -Z_{\theta c}/V_o \\ M_{uc} & M_{\gamma c} \end{bmatrix} \begin{bmatrix} \Delta u_c \\ \Delta \gamma_c \\ \Delta \theta_c \end{bmatrix} \quad (1)$$

The terms in the left hand side (lhs) matrix are made of airframe stability derivatives, augmented by feedbacks of the F matrix. The right hand side matrix is determined by the G matrix. The airplane can be made steady state decoupled by choosing G such that the rhs of equations (1) equals the lhs with $s = 0$. The lhs matrix determines the airplane's response dynamics; therefore, the F matrix may be used to obtain desired dynamics of the responses of the three variables to their controls, as well as to minimize coupling transients.

Analog computer traces showing the responses of the steady state decoupled (SSD) EBF STOL of Reference 2 are shown in Figure 4. It can be seen that this configuration has a minimal amount of transient coupling. The responses of the three variables to their controls are rapid and very well behaved. This situation was achieved by highly augmenting all stability derivatives with respect to the basic airframe values. High augmentation increases turbulence sensitivity. This will be discussed further in following sections.

Completely Decoupled (CD) Configuration - Transient, as well as steady state decoupling, results from the following set of equations:

$$\begin{bmatrix} (s - X_u) & 0 & 0 \\ 0 & (s + Z_\gamma/V_o) & 0 \\ 0 & 0 & (s^2 - M_\theta s - M_\theta) \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta \gamma \\ \Delta \theta \end{bmatrix} = \begin{bmatrix} -X_u & 0 & 0 \\ 0 & Z_\gamma/V_o & 0 \\ 0 & 0 & -M_\theta \end{bmatrix} \begin{bmatrix} \Delta u_c \\ \Delta \gamma_c \\ \Delta \theta_c \end{bmatrix} \quad (2)$$

In this configuration, feedbacks are used to null off-diagonal terms in the lhs matrix and, again, a prefilter, G, is used to obtain a diagonal right hand side matrix.

Unaugmented Navion derivatives were used in the left hand matrix diagonal with the exception of $M\dot{\theta}$ which was augmented by a factor of about two to improve pitch damping.

Analog computer responses to step inputs of this completely decoupled configuration are shown in Figure 5. As expected, they do not exhibit any coupling transients; also, all three responses are noticeably slower than those of the SSD configuration (Figure 4). The responses could have been made faster by augmenting the Navion derivatives in the diagonal of the left hand side matrix. However, this was not done as the responses were judged to be quite adequate and, therefore, there was no reason to bear the penalty of increased turbulence sensitivity associated with augmented stability derivatives.

"Recoupled" Controls (REC) - This configuration was identical to the completely decoupled (CD) one with one difference: pitch attitude response to control column was restored, such that column displacement caused a change in pitch attitude that was of equal magnitude and in the same direction as the flight path angle change produced by the control column in the completely decoupled configuration. The reason for including this configuration in the study was that attitude changes might provide the pilot with a good visual cue to predict variations in flight path angle. It was felt that the elimination of this cue in the constant attitude landings with the decoupled configurations might present some piloting problems and that they might be alleviated by recoupling pitch attitude to flight path angle as they are coupled in conventional aircraft.

EBF STOL with Conventional Stability Augmentation System (SAS) - This configuration was the subject STOL aircraft with its conventional SAS as tested in a previous flight simulation program that has been carried out at Princeton (Reference 5). The SAS included pitch attitude and pitch rate feedbacks to the horizontal tail to provide a pitch attitude command/hold system. Speed feedback to the flap and symmetric spoilers was incorporated in the system to enhance speed stability.

Improved SAS (ISAS) Configuration - A modified SAS configuration was evaluated in order to explore the possibility of improving handling qualities of the SAS airplane with a control system that is not as complex as the decoupling system. The modifications of the SAS were the following:

- Lift response to angle of attack, Z_{α} , augmentation was incorporated.
- The speed stability parameter, dY/du , was changed from zero to a small negative value.

- The thrust response lag was reduced (implying throttle to spoilers or flaps interconnect).
- The thrust control sensitivity, $Z\delta_{th}$, was increased.

Figure 6 shows a comparison of the SAS and ISAS configurations. The major advantage of the ISAS is an improved flight path response to control column input, having a faster rise time and settling down at a positive steady state value rather than going back to zero as in the SAS case. The Z_α augmentation is the main contributor to this improvement. The rise time of the flight path response to throttle was also shortened in the ISAS, mainly by the reduction of the thrust lag.

Simulation of Turbulence - Three components of turbulence were simulated in this program: fore and aft, u_g ; vertical, w_g ; and spanwise gradient, p_g . Turbulence signals that had been prerecorded on a magnetic tape were channeled into Navion controls to produce simulated turbulence-induced accelerations. The magnitudes of the accelerations were proportional to the appropriate stability derivatives of each configuration. Augmented, rather than bare airframe, stability derivatives were used in calculating turbulence sensitivity, as it was assumed that aerodynamic sensors were used in the feedback loops.

Lateral Directional Dynamics - Lateral directional dynamics that are typical for the subject airplane were employed in this program with all the various longitudinal configurations.

Adverse Ground Effect - A lift loss of $\Delta C_L/C_L \doteq -0.1$ was simulated. No moment or drag changes were included.

Test Procedure and Conditions

The test pattern that was used is shown in Figure 7. The approaches were either straight-in with 6° or 4° glide slope angles, or segmented with an initial 9° or 6° angle transitioning into a final 4° segment. The normal approach speed was 70 kts. The outer part of the approach was simulated IFR using a TALAR (Tactical Landing Approach Radar) MLS (Microwave Landing System) unit. At an altitude of 61 m (200 ft) the pilot transitioned to VFR, using optical glide slope light bars for guidance. The task required touching down within a 61 m (200 ft) long marked area on the runway with a rate of sink as low as possible.

Evaluation was mainly based on pilot rating using the familiar Cooper Harper scale adopted from Reference 6. Landing performance in terms of

touchdown distance and rate of sink was also measured. Most of the evaluations were flown by two Princeton University pilots.

RESULTS AND DISCUSSION

The Decoupled Configurations Compared to the STOL Airplane With Conventional SAS

General Aspects of Landing the Decoupled Airplanes - The steady state decoupled (SSD) STOL airplane turned out to be very well-behaved. The approach-and-landing was performed at a constant pitch attitude, and speed was constant at 70 kts. In calm air, speed and pitch attitude stayed at their selected trimmed values without any pilot intervention so that he had to control only flight path. Flight path response was judged to be very favorable in terms of quickness and authority.

A short period of adjustment was required for one of the pilots to accept the visual aspects of the constant attitude flare. The lack of aircraft rotation made this pilot underestimate the amount of flight path angle change associated with column deflection. This caused him to overflare. An overflare with the decoupled airplane resulted in a substantial touchdown overshoot since it maintained a parallel to the runway or slightly ascending flight path, and a constant speed unlike a conventional airplane that would lose speed and eventually settle down under similar circumstances. The decoupled airplane, following an overflare, had to be brought down by releasing pressure from the column. However, after a short learning period, this pilot was able to acquire the proper technique that resulted in a consistently precise touchdown at the middle of the prescribed zone, with a very low sink-rate and with a very small work load. The other Princeton pilot, having more experience with unconventional configurations, did not need even the short adjustment period. It turned out that the visual and normal acceleration cues that are available in a constant attitude flare, are quite adequate, enabling the pilot to land the airplane with remarkable precision.

Both pilots gave this configuration a Cooper rating of 2.0 for a 4° approach with no turbulence. This applies to the instrument part of the approach to the visual part, and to the flare-and-landing.

The completely decoupled (CD) configuration, in calm air, received ratings practically identical to those for the steady state decoupled (SSD) STOL, as might be predicted considering the similarity of the time responses of the two configurations.

Some quotations from pilot commentary about the decoupled configurations under no-turbulence conditions follow:

Pilot A: "This configuration, in the ideal no-wind, no-turbulence conditions tested, is very good. The airplane is controlled very easily on the glide slope where speed is maintained automatically and the pilot has to control γ only."

Pilot B: "Airplane can be landed smoothly and accurately; still, it is different from a conventional airplane in the fact that a normal airplane in the case of an overflare will eventually lose some speed and touchdown. In this airplane an overflare results in no touchdown unless pressure is released from yoke."

A flight path to column control sensitivity of $\Delta\gamma/\Delta\delta_c = 1.0$ deg/cm (2.7 deg/inch) was found to be a good value for the approach, and for a flare out of a 4° approach. A 35% higher value was preferred by the pilots for flying in simulated turbulence or landing out of a 6° approach.

The capability to trim out forces on the glide slope was considered an important feature of the decoupled system. A flight path trim rate of $\gamma_{TRIM} = 2^\circ/\text{sec}$ was found to be convenient. The trimmer was used quite extensively to make corrections on the approach.

6° Approach - No degradation in pilot rating was associated with the steeper, 6° , final segment of the approach. The pilots felt that the responsive control they had over flight path minimized the increase in work load that might have been associated with a higher rate of sink.

Ground Effect - Ground effect variations of $-0.1 < \Delta C_L/C_L < +0.20$ were tested and found to have no significant impact on landing the decoupled configurations. The pilots felt that with the tight flight path control that could be applied, they had no problem in counteracting the ground effect "suckdown." In contrast to these results, in the case of the STOL with conventional SAS, adverse ground effect contributed to piloting difficulties and accentuated other airplane deficiencies, as reported in Reference 5.

Figure 8 shows a landing time history for the steady state decoupled configuration for a 6° approach with no turbulence. It is very clear that the flare was performed by the column only. The pilot did not touch the throttle (or rather speed handle) throughout the flare. He made a small θ adjustment prior to flare initiation. Speed and attitude are essentially constant and $\Delta\delta_c$, $\Delta\gamma$, and h change smoothly from flare initiation to touchdown at $h = 0$.

Segmented Approaches - $6^\circ/4^\circ$ and $9^\circ/4^\circ$ approaches have been performed and have not been found to present any piloting difficulties with the decoupled configurations. The pilots felt that the task was quite easy to perform. The flight path trimmer was the main controller used by the pilots in turning the two segment corner. In calm air almost no control column motion was used and the corner was turned on trimmer only.

Aborted Approaches - Aborted approaches were tested with the decoupled configurations on several runs. The go-around technique for these airplanes was simply to pull on the column in order to command an up flight path angle. This indeed was very easy to do and presented no handling problem, and was rated at 2.0; however, the simulator climb performance was very limited in this maneuver. The Navion in this respect was probably not an accurate simulation of the STOL; however, the climb performance of the STOL is also very limited in the landing configuration, and therefore the simulation was not unrealistic. A rate of climb of 1.52 m/sec (300 ft/min) could be established, and the pilots considered it acceptable.

STOL with Conventional SAS - The STOL with conventional SAS was clearly a more difficult machine to control. It was described by the pilots as sloppy and sluggish and they had to work harder to obtain touchdown performance similar to that of the decoupled configurations. Landing it required the coordinated use of both column and throttle because neither alone provided adequate flight path control. The response of $\Delta\gamma$ to column contained a transient due to the presence of Z_w , but no steady state as $d\gamma/dv$ was zero. The transient was not big enough to permit using the column as the sole controller in the flare. The throttle, on the other hand, did produce steady state changes in γ , but because of the lag that was associated with it, the pilots could not use it for the rapid fine lift modulations that were required in the flare.

A landing time history for the SAS configuration at a 4° approach with no turbulence but with the nominal ground effect ($\Delta C_{L_{max}} = 0.1 C_{L_{\infty}}$) is shown in Figure 9. The extensive use of the throttle is in obvious contrast to the situation in the decoupled airplane. A throttle advance about 8 seconds prior to touchdown is employed in order to obtain the desired steady state change in γ .

Figure 10 provides a comparison of pilot ratings assigned to the completely decoupled (CD) configuration and to the SAS airplane. (CD and SSJ have the same ratings in the absence of turbulence.) With no turbulence the SAS is seen to be rated 3.0 for the approach, which is one unit worse than the CD configuration. For the landing, the SAS is rated 4.0, or two units worse than the decoupled airplane.

The improved SAS (ISAS) configuration is seen in Figure 10 to be rated significantly better than the airplane with the conventional SAS. With no turbulence it is rated only 0.5 units below the CD configuration for the approach and 0.75 units for the landing. The ISAS was found to be better than the SAS because of its improved flight path response to column and to throttle. However, the pilots still elected to use both hands in coordination while flying this airplane. They could rely more on the control column and less on the throttle in flaring the airplane, but they still did not have sufficient confidence to use a single control. This explains the slightly worse rating of the ISAS with respect to the decoupled configuration.

The Effects of Turbulence - Turbulence was simulated in this program as described in the Tested Configurations section. In the steady state decoupled configuration (SSD), as was indicated before, highly augmented stability derivatives were employed in order to achieve fast responses to pilot controls, and minimal coupling transients. Since it was assumed that aerodynamic, rather than inertial, sensors were used for augmentation, simulated turbulence disturbances were proportional to the augmented stability derivatives, rather than to those of the bare airframe. This resulted in an excessive sensitivity to turbulence of the SSD configuration that was beyond the control capability of the simulator. Therefore, this configuration had to be tested at a lower level of simulated turbulence (about one-fourth) than the other configurations. Using the same ground rules, the CD configuration was much less sensitive as its off diagonal stability derivatives were nulled (through the use of feedbacks), and all of its derivatives were lower than those of the SSD. This made it possible to test the CD configuration with the same amount of turbulence in which the conventionally augmented (SAS) STOL was tested; therefore, the CD, rather than SSD was used in comparing the effects of turbulence on the decoupled versus SAS configurations.

The main problem incurred by turbulence with all tested configurations, was heave upsets. Increased pilot effort was required to counter those upsets, depending on their magnitude and on the effectiveness of the available flight path control. Speed disturbances were also caused by turbulence; however, since they were mostly of low magnitude and frequency higher than available with speed control, the pilots chose to accept those disturbances rather than try to fight them.

Figure 11 provides a comparison of landing the decoupled configurations with and without turbulence. The turbulence related heave motions can be seen on the ΔY trace, and the increased pilot activity is very obvious in the control column trace.

The effect of turbulence on pilot ratings of the various configurations can be seen in Figure 10. The CD in landing is seen to be degraded by turbulence from a rating of 2.0 to 3.0 which is still a good position in the Cooper scale. The SAS is degraded from 4.0 to 5.5 which is not too good, as a rating higher than 6.5 puts an airplane under the "major deficiencies" label. The ISAS in landing with turbulence is rated at 4.0, which is not quite as good as the decoupled airplane, but it is significantly better than the SAS.

It should be emphasized that a high sensitivity to turbulence is not inherent in the steady state decoupling concept. The particular configuration that was tested in this study turned out to be that way since, as discussed previously, the best possible transient response was sought in the design, whereas turbulence sensitivity was not taken into account. Turbulence sensitivity is an important factor affecting aircraft handling qualities and it should be taken into account in the design of an augmentation system. Turbulence sensitivity may be reduced by:

- Avoiding highly augmented stability derivatives if aerodynamic sensors are to be used.
- Using inertial sensors.
- Filtering.

Indeed, the turbulence sensitivity of the SSD configuration has been reduced in the moving base simulator study of Reference 3.

Landing Performance

Figure 12 compares the landing performance obtained with the decoupled airplane to the performance of the conventionally augmented STOL. The data used in this figure included runs with turbulence and without it. The median touchdown point of both configurations is seen to be very close to the center of the touchdown zone (0 in the figure). The spread is very small for the decoupled airplane and somewhat bigger for the SAS. The median rate of sink of the decoupled control scheme is a very good deal less than 0.3 m/sec (1 fps), and its spread is very small. The median rate of sink of the SAS is significantly higher, at 0.6 m/sec (2 fps), but still acceptable. The SAS rate of sink spread is markedly higher than the results with the decoupled configuration.

Figure 13 shows a comparison of results obtained with the decoupled airplanes in the in-flight simulation versus the ground simulator studies of References 2 and 3. It is very obvious that the in-flight results are much

better than those of both fixed base and moving base ground simulators. Going from fixed base to moving base ground simulator, seems to improve noticeably the rate of sink performance, but hardly affects touchdown distance results.

The Recoupled Configuration

The recoupled configuration was derived from the completely decoupled one by restoring pitch attitude response to control column, equal to the flight path angle change, as described in the Tested Configurations section. This was done in an attempt to evaluate the role of pitch attitude as a cue that might help the pilot in judging flight path changes. This configuration was introduced into the program after the pilots had flown the decoupled airplanes extensively. The surprising result was that the pilots found the recoupled control scheme to be somewhat less desirable than the decoupled one. The first reaction of both pilots to landing the recoupled airplane was to rate it by 0.5 unit worse than the decoupled case. After some more experience had been gained, one of the pilots thought that the recoupled configuration might be equivalent to the decoupled one, whereas the other pilot retained his opinion that recoupling caused some degradation of flying qualities.

Pilot commentary: "No advantage over the decoupled airplane. ILS tracking: attitude coupling neither helps nor hurts. Visual tracking: don't get much help out of attitude change. Get information from [the glide slope] lights. There is a problem for close-in down-gamma correction; I don't like the nose going down in response to a down-gamma command. Flare and touchdown: very much like a normal airplane. Nose up and down motion interferes with good prediction of the touchdown point. The ability to judge sink rate close-in is not improved by the theta coupling."

To conclude, the results of the tests indicate that coupling attitude to flight path does not provide any advantage over the decoupled situation. This suggests that the perceptions obtained by the pilot from the visual and motion cues that are available in a constant attitude landing, are clearly adequate for the task and no additional improvement is obtained by using pitch attitude changes.

Deceleration Prior to Touchdown

It was felt that it might be advantageous for the powered lift STOL airplane to approach at a speed that is somewhat higher than the touchdown speed, as this might improve handling qualities. and performance for a

possible aborted approach. This, of course, necessitates a speed reduction prior to touchdown. Runs in which the pilots were requested to fly the approach at 75 kts and touchdown at 70 kts were incorporated in the flight program in order to evaluate the effect of the deceleration maneuver on piloting with the various configurations. Speed reduction was performed with the decoupled configurations by simply pulling back on the speed command lever (throttle). In the ISAS configuration, a simulated autopilot speed command dial was employed, and in the SAS, coordinated control column and throttle movements had to be used in slowing the airplane down. The pilots were requested to reduce speed below breakthrough altitude, but no precise point was specified. It turned out that an increment in workload was caused by the deceleration in all cases. The amount of this increment depended on how well the approach was going prior to the deceleration, and on the configuration. With the decoupled configurations deceleration was easy and could be done in coincidence with the flare. The ISAS and SAS configurations required an ever increasing effort. The increased difficulty in performing the deceleration was offset by the pilots' choice to slow the harder configurations down at a higher altitude.

Further details on this subject are given in Reference 4.

CONCLUSIONS

Three variations of decoupled longitudinal control configurations and two variations of conventionally augmented longitudinal control configurations for an EBF STOL airplane in the approach and landing flight phases were studied through in-flight simulation. The conclusions of this study follow:

1. The decoupled longitudinal controls that were tested produced very favorable flying qualities in the approach and landing. This resulted in small touchdown dispersions along with consistently low sink rates.
2. Adverse ground effect did not cause any piloting problems with the decoupled airplane as it could be easily countered by control actions.
3. Segmented approaches, in which an initial 9° glide slope segment was followed by a shallower 4° or 6° final segment, were easy to perform with the decoupled airplane. Turning the corner between segments was found to be an easy task. No significant degradation was associated with using a 6° rather than a 4° final glide slope.

4. No significant differences in flying qualities were found between the steady state and completely decoupled control concepts. This conclusion would seemingly apply for any steady state decoupling scheme in which transients were kept sufficiently small and of short duration.
5. The extreme turbulence sensitivity of the steady state decoupled control scheme that was tested in this study, resulted from the use of high feedback gains along with aerodynamic sensors as sources of feedback signals. High sensitivity to turbulence is not inherent in the steady state decoupling concept. It may be reduced by either using inertial sensors, or by avoidance of high gains in feedback paths. Sensitivity to turbulence must be considered in the design of such a control system.
6. The completely decoupled configuration which, under the assumptions adopted in this study, had a low sensitivity to turbulence produced in simulated turbulence a very significant improvement in flying qualities with respect to the conventional SAS that was tested.
7. The recoupled configuration in which pitch attitude changed along with flight path, but which was otherwise decoupled, did not result in any improvement of flying qualities with respect to the decoupled configurations. This suggests that the perceptions obtained by the pilot from the visual and motion cues that are available in a constant attitude landing, are clearly adequate for the task and no additional improvement is obtained by using pitch attitude changes.
8. The conventionally augmented airplane that was tested had significantly poorer handling qualities than the decoupled configurations, especially for landing in turbulence and with adverse ground effect. The piloting problems of this airplane resulted mainly from its low lift response to angle of attack, the lag associated with the throttle response, and the resulting necessity to use coordinated control column and throttle action in controlling flight path.
9. Modifying the conventional SAS by increasing lift response to angle of attack, Z_{α} , changing $d\gamma/du$ from zero to a small negative value, increasing thrust control sensitivity, Z_{δ_t} , and reducing thrust response lag, yielded a significant improvement in pilot rating, but the overall level of handling was judged to be not as good as for the decoupled configurations.

REFERENCES

1. Grantham, W. D., Nguyen, L. T., Patton, J. M., Jr., Deal, P. L., Champine, R. A., and Carter, C. R.: Fixed-Base Simulator Study of an Externally Blown Flap STOL Transport Airplane During Approach and Landing. NASA TN D-6898, October 1972.
2. Miller, G. K., Jr., Deal, P. L., and Champine, R. A.: Fixed Base Simulation Study of Decoupled Controls During Approach and Landing of a STOL Transport Airplane. NASA TN D-7363, February 1974.
3. Miller, G. K., Jr. and Deal, P. L.: Moving-Base Visual Simulation Study of Decoupled Controls During Approach and Landing of a STOL Transport Aircraft. NASA TN D-7790, January 1975.
4. Feinreich, B.: In-Flight Simulation Study of Decoupled Longitudinal Controls for the Approach and Landing of a STOL Aircraft. NASA CR-2710, to be published summer 1976.
5. Ellis, D. R.: An In-Flight Simulation of Approach and Landing of a STOL Transport with Adverse Ground Effect. Princeton University Report No. 1267, December 1975.
6. Harper, R. P., Jr. and Cooper, G. E.: The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities. NASA TN D-5153, April 1969.

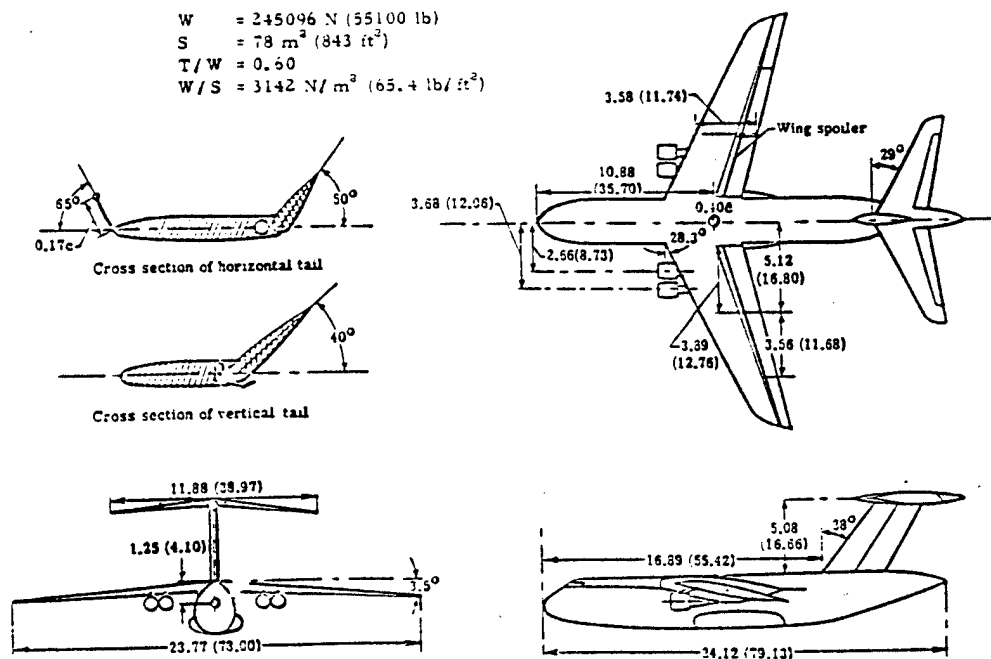


Figure 1. Three-View Drawing of Simulated Airplane (All Linear Dimensions are in Meters (ft))



Figure 2. The Princeton In-Flight Simulator

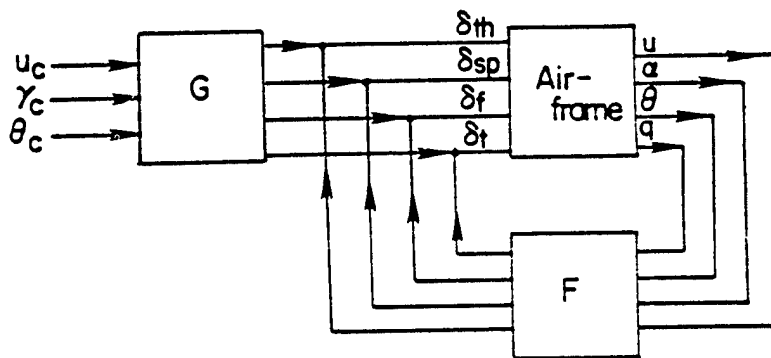


Figure 3. Schematic Block Diagram of the Steady State Decoupling System

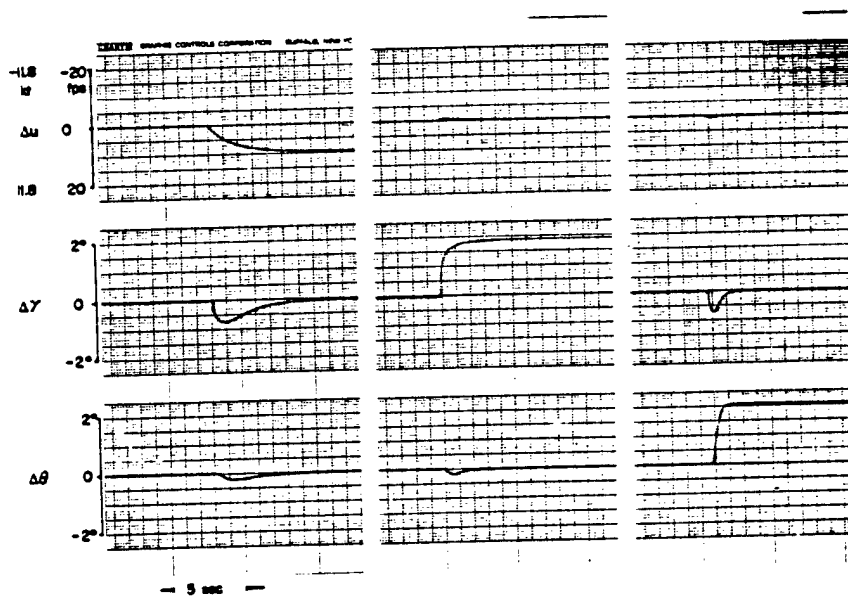


Figure 4. Analog Computer Generated Responses of the Steady State Decoupled Configuration to Step Inputs

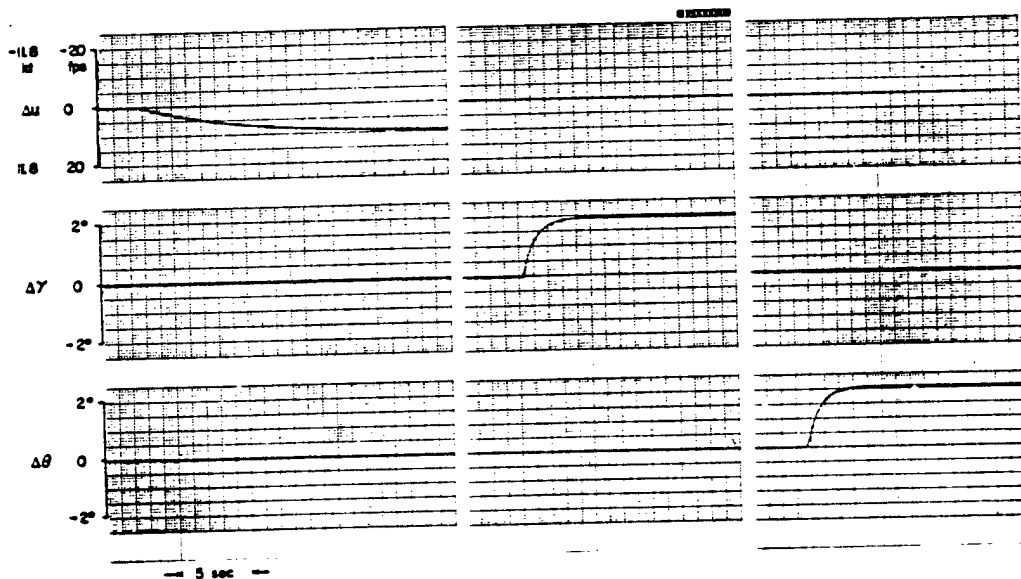


Figure 5. Analog Computer Generated Responses of the Completely Decoupled Configuration to Step Inputs

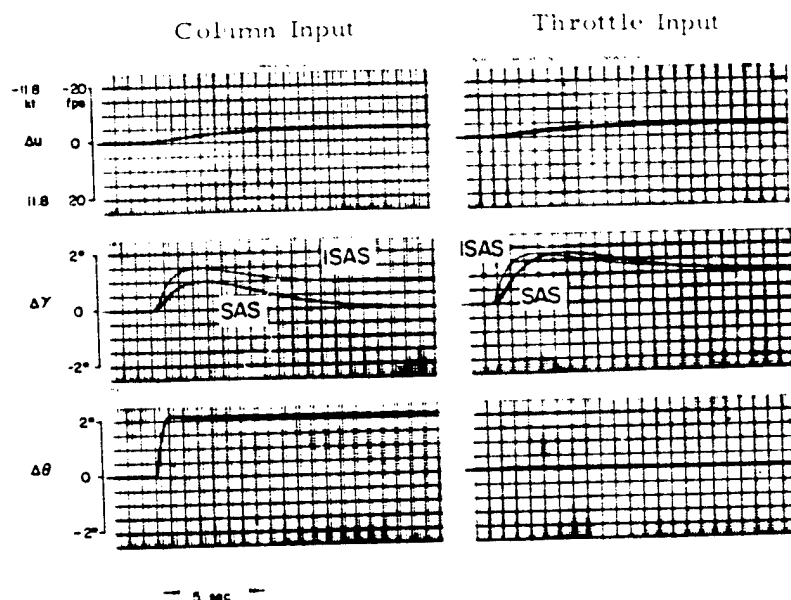


Figure 6. Analog Computer Traces of the SAS and ISAS Configurations' Responses to Step Inputs

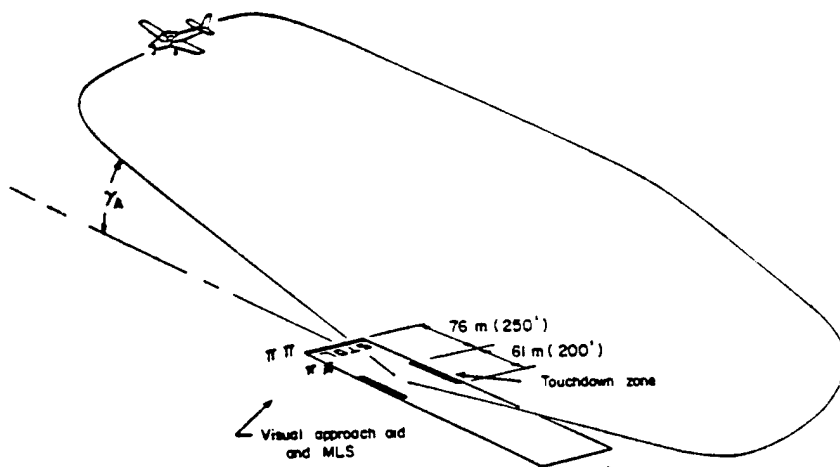


Figure 7. Simulated STOL Runway and Flight Pattern

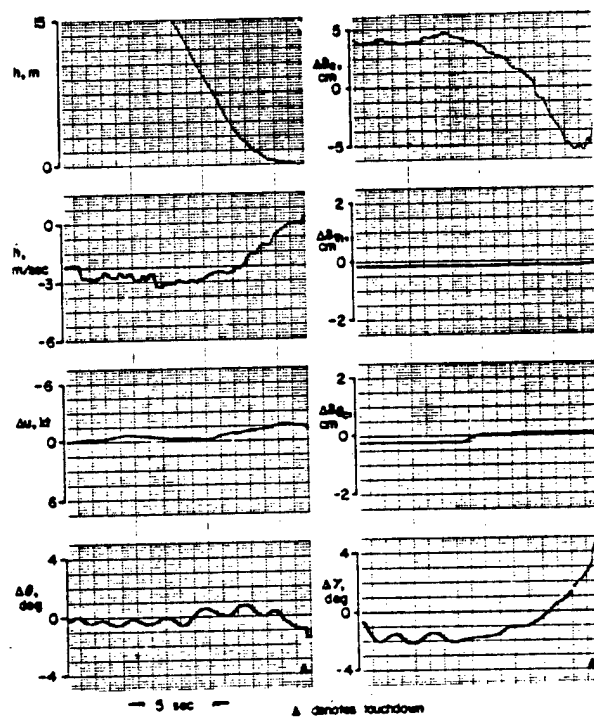


Figure 8. Landing Time History of the Steady State Decoupled Configuration.
6° Approach; Adverse Ground Effect, No Turbulence.

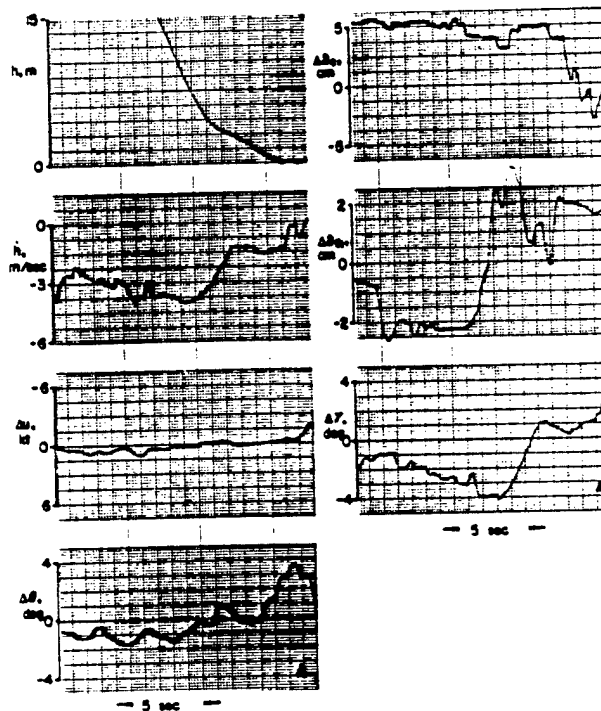


Figure 9. Landing Time History of the SAS Configuration. 4° Approach; Adverse Ground Effect; No Turbulence.

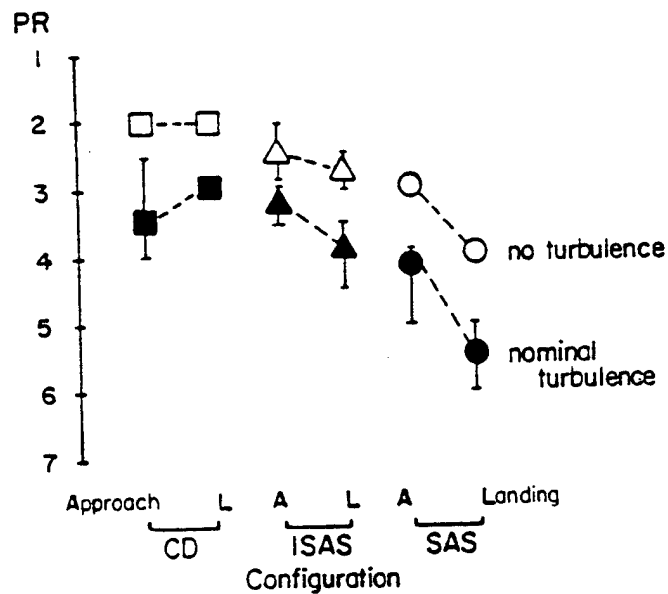


Figure 10. Pilot Ratings for the Approach and Landing of the CD, ISAS, and SAS Configurations

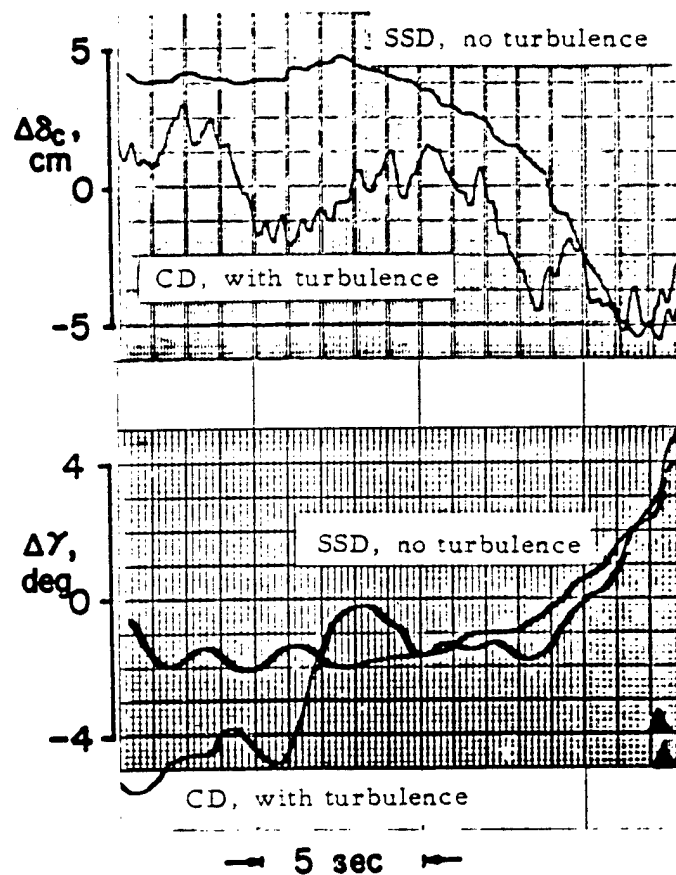


Figure 11. Superimposed Landing Time Histories of CD Configuration with Turbulence, and the SSD with No Turbulence

ORIGINAL PAGE IS
OF POOR QUALITY

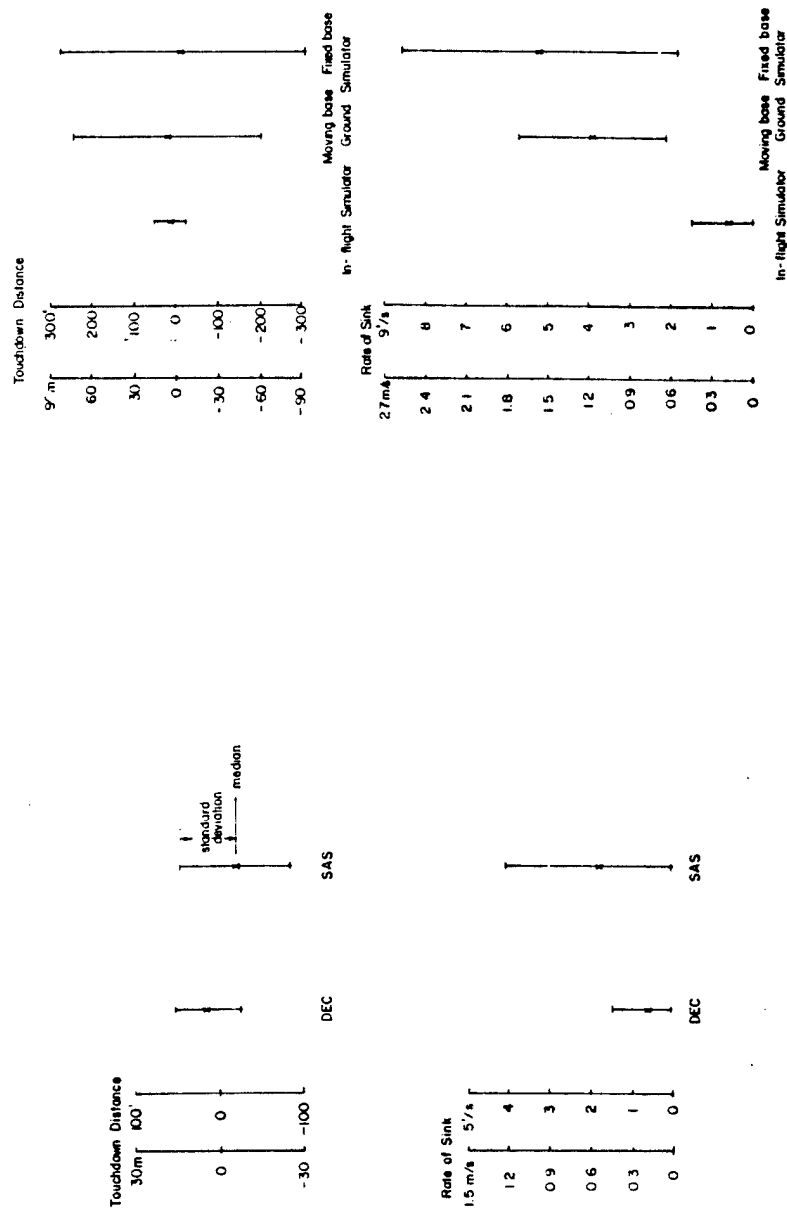


Figure 12. Landing Performance Results

Figure 13. Landing Performance of the Decoupled Airplane Flight vs Ground Simulators

HIGH ACCELERATION COCKPIT DISPLAYS AND CONTROLS

By James A. Townsend and Norbert J. Kropewnicki

Air Force Flight Dynamics Laboratory

SUMMARY

Modern fighter maneuverability has outstripped the pilot's physical capability to perform in air-to-air combat. The objective of the Air Force High Acceleration Cockpit program is to provide pilots with increased physiological protection during sustained and repeated high G air-to-air combat maneuvers. A major breakthrough in useful combat maneuverability can be achieved by providing the pilot with an advanced cockpit design incorporating an articulating pilot seat which allows the pilot to reposition himself for high energy, high G maneuvering. Controls, displays, ejection seat, restraining system, seat supinating, reclining actuation system, and other man-machine interfaces are included in the cockpit design.

This paper describes the High Acceleration Cockpit (HAC) Program in general with emphasis on the fighter cockpit design relationships, including side-stick and throttle characteristics and the design of displays to retain total mission capabilities of the aircraft. Results of pilot evaluations of complete mockups and highlights of other work to date are presented.

INTRODUCTION

A new concept in the field of fighter aircraft crew station design and control display technology has emerged in the form of the High Acceleration Cockpit, commonly referred to as the HAC cockpit. This concept has been developed since 1970 and is just now in the threshold of a full scale evaluation in a TF-15 fighter aircraft. The latest complete step to date in the development of this concept is the HAC Applications Study - the basis of this paper.

BACKGROUND

The beginning of the development activity in HAC was founded on the growing appreciation of the fact that new fighter designs were outgrowing the human pilot's capabilities in sustained high-G (3 to 7) turns

repeated several times in an air combat engagement. Not only were the pilot's maximum G tolerances going to be good over too short a time span compared to the airplane's continuous high G maneuvering capability, but the fatigue effect of repetitive exposures would dramatically reduce post-G tracking accuracy and ability to fly several combat missions a day in a wartime scenario.

The HAC concept, that of repositioning the pilot transverse to the load vector is intended to significantly increase his G tolerance, reduce his workload and fatigue, and improve his cognitive capability.

Figure 1 illustrates the effect of HAC seat back angle on increasing the useable ACM envelope.

Figure 2 illustrates the principle of the HAC seating concept, which provides a reduction in the height of the hydrostatic column between eye and heart by allowing the G forces to act transversely. Blood pressure at eye level can be maintained and blood pooling in the lower extremities can be reduced with a consequent lowering of heart rate.

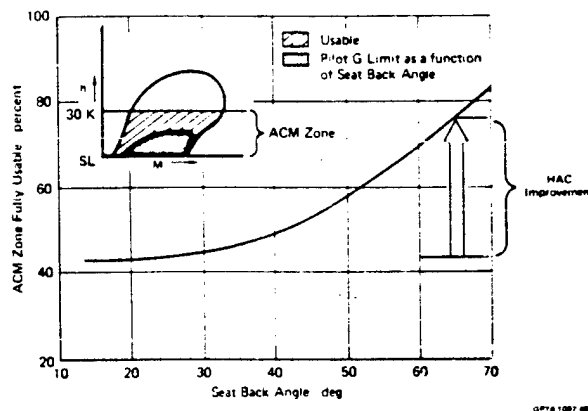


Figure 1
Usable ACM Envelope Expansion
For Continuous Sustained Maneuvering
on a Routine Basis

Centrifuge tests conducted at Wright-Patterson AFB, (AMRL) and Brooks AFB, (AMD/SAM) (Ref. 1) have demonstrated that the HAC concept provides significant improvement in pilot cognitive functioning and physiological condition in the 3 to 10 G range when the seat back angle is rotated above 45 degrees and up to 65 degrees. A 2 to 3 G improvement in normal G tolerance has been shown with substantial increase in tracking performance and comfort at all G levels during combat type maneuvers.

Background studies in the last two years have also included two manned air combat simulations at McDonnell Aircraft Company's dome simulator. Manned simulation involved subjects from Tactical Air Command. Results indicated a distinct and consistent combat advantage for the HAC equipped "aircraft" and a reduction in overall high G exposure because of the early advantage attained over the non-HAC equipped adversary. (Ref. 2).

Indications from the latest of these simulations based on one-on-one encounters comparing a HAC-equipped 10.5 G fighter with a non-HAC equipped 7.5 G fighter of otherwise similar design against a common threat aircraft include:

- a. 20% more first offensive opportunities for HAC A/C (Engagements converted to first kill).
- b. 20% reduction in threat a/c first conversion capability.
- c. 50% increase in cumulative missile hits in the first 40 seconds of the encounter.
- d. 200% increase in cumulative bullet hits in the first 40 seconds of encounter.
- e. 80% increase in kill ratio in the first 40 seconds of encounter.
- f. 75% reduction in load limit exposure of HAC fighter versus non-HAC fighter.
- g. Averag. G level reduced from 3.5 G for non-HAC a/c to 3.3 G for HAC a/c.
- h. Improved tracking
- i. Improved vision
- j. Reduced pilot workload-physical and cockpit operations.

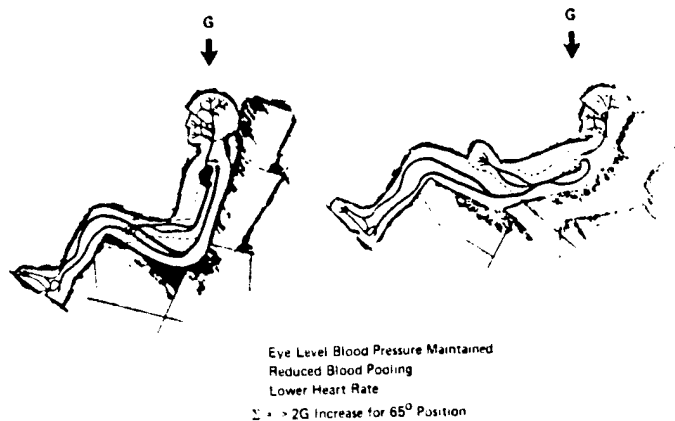


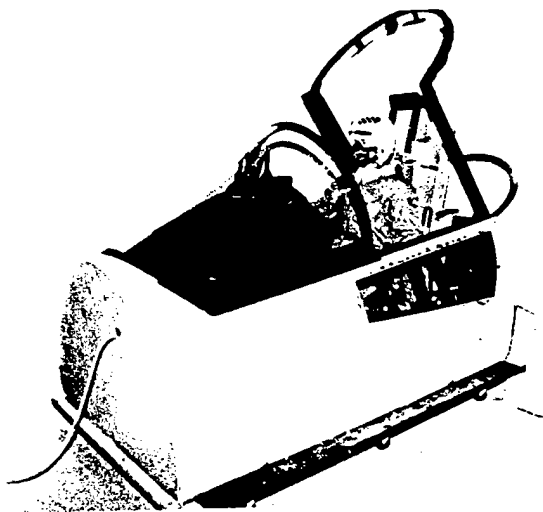
Figure 2

THE APPLICATIONS MOCK-UP STUDY

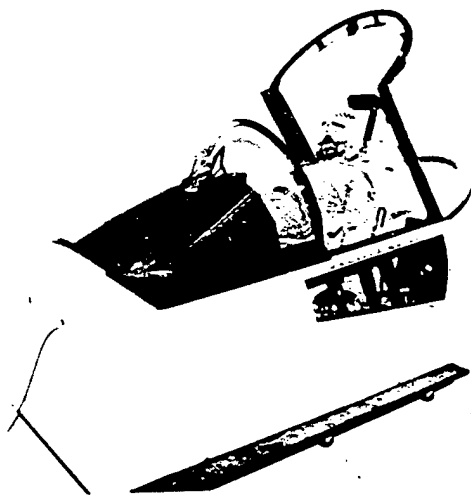
Our mock-up or design aid study consisted of the following major elements (Ref. 3):

- a. Construction of a full scale cockpit mock-up of the single seat version of the F-15.
- b. Modification of a conventional ejection seat to include the HAC reclining/lifting capability (positioning system).
- c. The effect of the pilot positioning system on the pilot's ability to accomplish his mission tasks.
- d. Flight controller (side arm) and throttle integration and evaluation for the range of seat positions.
- e. Evaluation of control/display concepts designed to provide accessibility and visibility from all seat positions.
- f. Test Plan preparation and static evaluation of candidate cockpit configurations by six operational fighter pilots.

Figure 3 depicts the HAC concept in the full scale mock-up design aid. The standard ejection seat and launch rails are in the normal upright position. The HAC articulating seat liner can be raised to the reclined and lifted position at the pilot's option when he needs load factor protection. The cockpit was designed so that while flying



Upright



Reclined

Figure 3.- HAC F-15 Cockpit Mock-up

hard maneuvers he retains outside-the-cockpit vision, can see necessary inside-the-cockpit displays, and can reach the required controls and switches. He is provided with side arm controller and low-throw rudder pedals both of which are manually adjustable for the range of 5th to the 95th percentile pilots. Partial automatic adjustment with seat articulation is also provided.

EVALUATION OF THE DESIGN AID

The pilot evaluators were subjected to a series of tests including anthropometric measures, vision envelopes, task performance times and eye and head movements.

Pilot tasks during each simulated mission phase received subjective comments. The crew also ranked individual cockpit items, priority of use, and design location by a "Paired Comparison" technique. Pilot comments were obtained with regard to: control and display design, grouping and location, seat adjustment, back angle, body support, restraint, escape system, and external and internal vision. Two over-all display and control arrangements were presented and evaluated: Configuration A, Figure 4, and Configuration B, Figure 5. The first represented off-the-shelf components suitable for use in a near term flight test evaluation, and the second a concept depicting an advanced HAC cockpit using technology compatible with the 1980/85 time frame. This latter version, Configuration B, maintained a control/display capability equivalent to the baseline F-15 aircraft.

Considerable effort was expended and several iterations were made in finalizing the design and conducting the evaluation of the primary flight and propulsion controls, side arm and throttle (Figure 6). A previous study had also investigated this area and evaluated other controller locations (Ref 4).

The pilot subjects felt that the concept was good and would provide a very real advantage in the combat area. The HAC seat with minor improvements was considered as satisfactory for follow-on testing in an in-flight program. Pilot tasks, visibility and reach, and controls and displays were judged acceptable for an actual aircraft installation and evaluation. Actual production application if decided upon would require Configuration B, as the near term Configuration A was considered compromised for full mission compatibility.

Areas of concern were mainly those which require in-flight evaluation. Control capability with the sidearm controller will require close attention, since precise tracking is required in the fighter mission. High fidelity simulation will be used for this component prior to flight tests. Further effort is also required in design and ground tests to obtain a flightworthy ejection seat design with its articulating mechanism prior to flight.

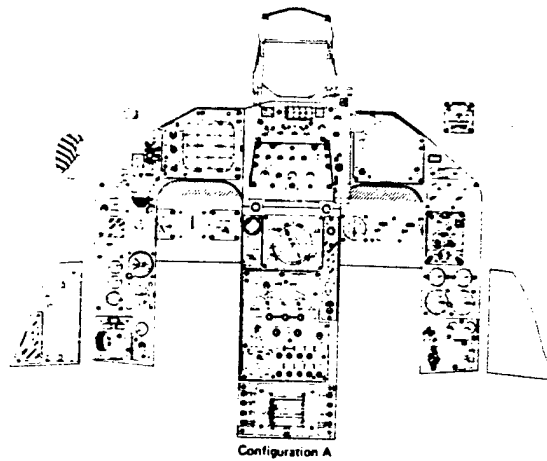


Figure 4

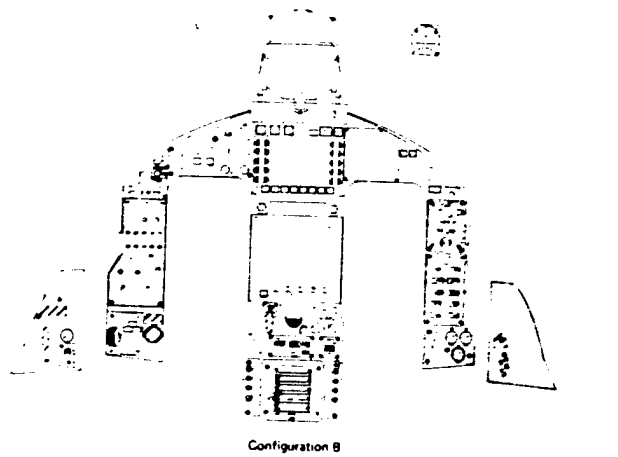
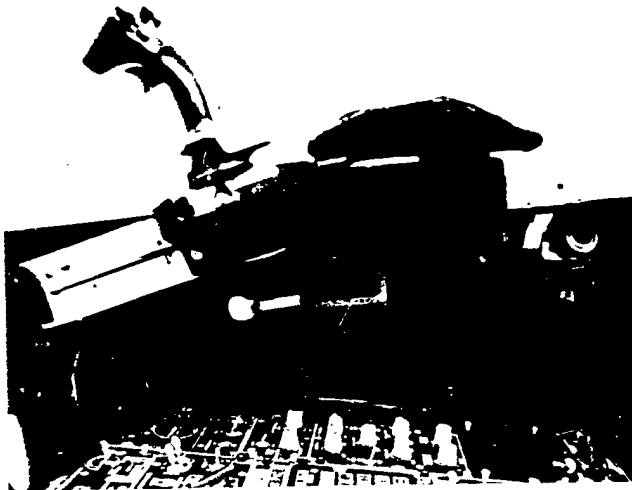
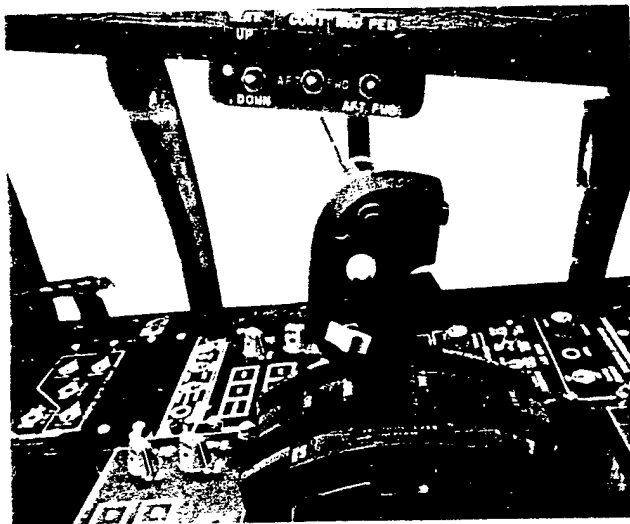


Figure 5

ORIGINAL PAGE IS
OF POOR QUALITY



Side Arm Controller



Throttle

Figure 6

A basic issue emerging from the evaluation pointed out a requirement to develop multi-sensor displays for integration into the cockpit if the HAC concept is ever to become a production system. The need for these displays is a result of the reduced main instrument panel area due to the interference of the pilot's legs in the reclined position. Several programs, independent of HAC, dealing with such displays are currently in the development phase at Wright-Patterson AF Base. An example is the Multimode Matrix Display (MMM), which is a short depth, light weight, low power level device using Light Emitting Diodes (LED) to display primarily alpha-numerics.

FUTURE PLANS

The next plateau in the overall High Acceleration Cockpit program is the design, fabrication, and flight test of the cockpit in a TF-15 aircraft. This phase is about to commence in the form of an Advanced Development Program (ADP). The flight test will include basic tracking maneuvers at various G levels and seat back angles, simulated combat with an adversary aircraft, and demonstration flights for interested service and industry personnel. The ADP contract, with McDonnell Aircraft Company, will require 30 months. Flight tests will be run during the last half of 1978.

REFERENCES

1. Burns, J. W.: A Re-evaluation of a Tilt-Backseat as a Means of Increasing Acceleration Tolerance. USAF SAM/VNB, 1974, Brooks AFB, Texas.
2. Sinnett, J. M., Prouhet, E. P., et al: High Acceleration Cockpit Simulator Evaluation. AMRL-TR-75-123, 1976, in draft.
3. Mattes, R. E., et al: High Acceleration Cockpit Applications Study. AFFDL-TR-75-139, 1976, Volumes I to IV.
4. Mattes, R. E., Asiala, C. F.; High Acceleration Cockpit Controller Locations. AFFDL-TR-74-48, 1974, Volume I - Program Summary.

SESSION IV
NONVISUAL DISPLAYS

Chairman: W. H. Levison

PRECEDING PAGE BLANK. NOT FILMED

DESIGN OF AN ENGLISH LANGUAGE DATA

BASE QUESTION-ANSWERING SYSTEM

By David L. Waltz

University of Illinois
Coordinated Science Laboratory and
Department of Electrical Engineering
Urbana, Illinois 61801

SUMMARY

A question-answering system, PLANES, has been written to allow a casual user to obtain information from a large data base by typing natural English questions to the system. This paper discusses design goals, gives examples of PLANES' operation, and discusses helpful factors and design difficulties in the PLANES world.

INTRODUCTION

A prime obstacle for non-technical people who wish to use computers has been the need to either learn a special language for communicating with the machine or communicate via an intermediary. We feel that the time is ripe for computers to be equipped for natural language systems which can be used by persons who are not trained in any special computer language. In order for such systems to be of value to a casual user, the systems must tolerate simple errors, must embody a degree of "common sense," must have a relatively large and complete vocabulary for the subject matter to be treated, must accept a wide range of grammatical constructions, and of course must be capable of providing the information and computations requested by the user.

We are developing such a system called PLANES (for Programmed LAnguage-Based Enquiry System) at the University of Illinois Coordinated Science Laboratory (see reference 1). PLANES includes an English language front end with the ability to understand and explicitly answer a user's requests and to carry on clarifying dialogues with him, as well as the ability to answer vague or poorly defined questions. We are also building a library of associated programs which includes functions for recognizing patterns within the data base and for alerting a user when certain patterns of data occur which are of interest to him. This work is being carried out using a subset of the U.S. Navy 3-M data base of aircraft maintenance and flight data, although the ideas can be directly applied to other record-based data bases, both military and non-military.

THE GOAL OF PLANES

Our main goal is to allow a non-programmer to obtain information from a large data base with no prior training or experience. A system to realize this goal (1) must be able to understand to a substantial degree a user's natural language and (2) must be able to help guide and educate the user to formulate requests in a form that the system can understand (see reference 2).

Subgoals of PLANES

We have formulated a number of subgoals which we feel are important for realizing our main goal:

(1) The system must accept a user's "natural" English input, possibly including complex syntactic constructions, abbreviations, pronoun reference, and ellipsis (i.e. omission of one or more words that can be obviously understood in context).

(2) The system must provide explicit answers to questions, and not merely retrieve a file which somewhere contains the answer. The system should phrase its answer in a clear manner, including units or dimensions of numerical answers and a description of the answer values. Whenever possible, graphical output is desirable.

(3) The system must be tolerant of minor errors, e.g. spelling and grammatical errors; it should suggest corrections for user approval whenever possible, and should in general be able to continue processing of the corrected request without requiring a complete retyping of the request.

(4) The system should use clarifying dialogues for several purposes:

(a) to feed back its understanding of the user's request, so that the user can feel confident that the system has understood his request.

(b) to ask the user pointed questions about portions of a request which it does not understand, in the hope of evoking a paraphrase it can understand.

(c) to add new words, phrases and sentences to the system's knowledge base.

(d) to provide appropriate HELP file information in the event of user errors or direct requests for help.

(e) to provide information about the system's capabilities, abbreviations it knows, general contents of the data base, and other such information to help orient a new user.

(5) Such a system should be convenient to use:

(a) it should be interactive and on-line,

(b) it should operate rapidly. One minute seems to us to be a critical length of time: if one in general gets responses in less than a minute, one retains a level of involvement and interest in the interactive process; past a minute, interest begins to wane, and boredom or impatience sets in.

(c) the system should require a minimum of typing; abbreviations and ellipsis should be routinely handled.

(6) The system should be relatively easy to extend, both within its own world, and to new data bases and domains of discourse.

It is unlikely that any design can simultaneously satisfy all these subgoals in an optimum manner. In fact, in PLANES we have chosen a solution which is not particularly easy to extend, either within our world or to another world. In addition, in our current implementation we have not included mechanisms for satisfying all the clarifying dialogue subgoals. In particular, the system now fails totally if it does not recognize any one of the phrases in a sentence. Unlike the problems of extending the system, this subgoal can be realized within our model relatively easily. We have achieved considerable success on all other subgoals, though our current system (as of May 1976) will answer probably no more than 25% of the questions phrased by a novice user.

EXAMPLES OF PLANES' OPERATION

All examples in this section are taken from actual transcripts of PLANES' operation. Figure 1 is a complete transcript of the operation of the current planes system on a single question. Note that the paraphrase and answer are expressed in a form which is not as intelligible as an English form would be. We intend to soon augment the system to provide English paraphrase and answer forms.

Figure 2 shows an extended session, with all but the question and answer suppressed. Figure 2 illustrates the ability of the system to handle ellipsis (i.e. information omitted which is to be understood in context).

Figure 3 shows PLANES' ability to provide answers in a graphical form. NORMRS stands for "Not Operationally Ready HouRS"; ACTDATE stands for "ACTION DATE"; 2001 is the first day of the year 1972.

Figure 4 shows a short dialogue. PLANES first asks for time period information necessary for performing its data base search. It then detects a misspelling and suggests two plausible corrections to the user. Once the user selects the appropriate correction by typing "2", the system is able to continue processing without requiring further typing.

Please enter your question.....

HOW MANY FLIGHTS HOURS DID PLANE 0003 LOG IN JAN 73

parsing.....

{cpu time was 1.74 seconds, real time was 5.95 seconds.

I have understood your request as:

(COUNT TIME

(NET NIL)

(FLY (PLANE (PRONOUN NIL)

(TYPE PLANE)

(BUSER 3.)

(PLNEG NIL)

(PLDAM NIL)

(PLMAI NIL))

NIL)

(TIME (DATE (MONTH (1. 0. 0.))

(DAY NIL) (YEAR 73.)) NIL))

Interpreting.....

{cpu time was 0.29 seconds, real time was 3.71 seconds.

I have interpreted your request as follows:

(FIND ALL

((V 0))

((SUM (V TOTHR))

(AND (EQU (V ACTDATE) 301.) (EQU (V BUSER) 3.))

NIL)

Evaluating.....

{cpu time was 6.29 seconds, real time was 16.9 seconds.

((SUM TOTHR) = 33.)

Figure 1. Complete transcript of PLANES' answer
to a user query

>> How many flights did the A7 with tail number 003 make in
January, 1973?

((SUM TOTFLTS) = 17.)

>> How many flights did plane 003 make in Feb. 73?

((SUM TOTFLTS) = 1.)

>> During April?

((SUM TOTFLTS) = 8.)

>> March?

((SUM TOTFLTS) = 13.)

>> All of 1973?

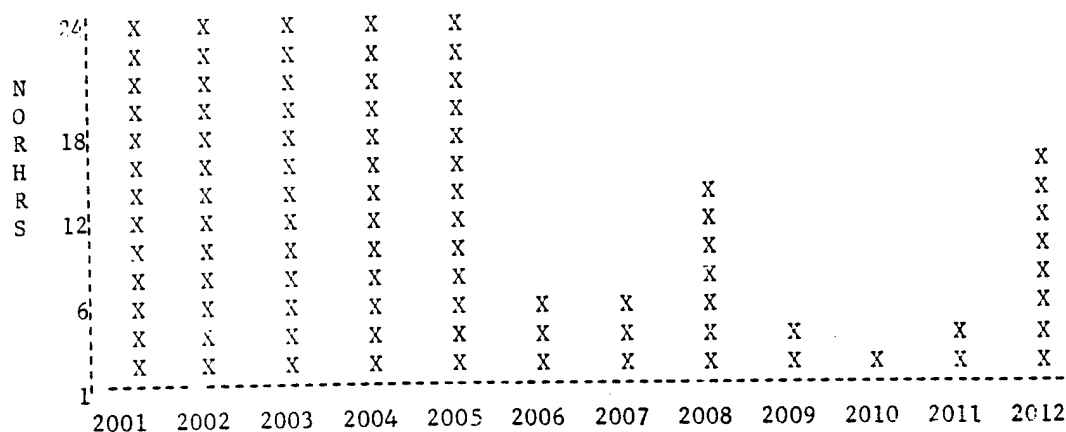
((SUM TOTFLTS) = 39.)

Figure 2. Partial transcript of dialogue showing
PLANES' handling of ellipsis

>> plot how much maintenance occurred to plane 2 between Jan. 1 and Jan. 12, '72.

Evaluating.....

{cpu time was 29.32 seconds, real time was 38.11 seconds.



ACTDATE

MIN = 1.0

AVERAGE = 14.0

MAX = 24.0

Figure 3. PLANES' graphical presentation of an answer

>> Which Skyhawks logged less than 10 hours

PLEASE ENTER A SPECIFIC TIME PERIOD:

>> between Jan. 1 and Jule 1 '73.

I DON'T KNOW THE MEANING OF JULE
PERHAPS IT'S MISSPELLED?

PLEASE ENTER A NUMBER CORRESPONDING TO ONE OF THE FOLLOWING:

1. JUNE
2. JULY
3. NONE OF THE ABOVE

>> 2.

OK

<Answers question>

Figure 4. Example of PLANES' dialogue
and spelling correction abilities

Figure 5 shows PLANES' ability to answer general questions about its contents.

THE PLANES WORLD

In this section I describe the general environment in which the language understanding programs operate. This environment includes the data base, the user, and the user's range of queries.

The PLANES Data Base

We have obtained a data base from the Navy 3-M Data Base for Aircraft, Mechanicsburg, PA., consisting of complete records of aircraft maintenance and flight information for 48 A7 and F4 aircraft, extending over a period of two years. Each time a plane is serviced, a record is made including such information as the time and duration of the maintenance, who performed it, what action was taken, which parts were used, the manufacturers of these parts, whether or not the service was scheduled or unscheduled, and so on. Records on the number of flights and the number of hours in the air are also kept for each plane. There are roughly forty different record formats which occur in the data base, each containing between ten and twenty separate fields, where each field encodes information like the date of the action, type of aircraft, serial number of the aircraft, type of malfunction, component serviced, the work station performing maintenance and so on.

The 48 aircraft in our data base are divided into three groups:

- (1) 24 planes which crashed or sustained major damage in accidents involving mechanical failures;
- (2) 12 planes with bad maintenance records; and
- (3) 12 planes with good maintenance records.

"Good" and "bad" records were judged by comparing the ratio of the number of NOR (Not Operationally Ready) hours to the number of flight hours. A high ratio represents a bad record while a low ratio corresponds to a good record.

In addition, we have summaries of maintenance and flight data for all F4 and A7 aircraft for the same two year period, so that we can have some basis for classifying events as "normal" or "unusual."

The PLANES data base contains on the order of 10^8 bits, and occupies about one third of a DEC RPO4 disc pack in compressed form. This data base, while quite large, represents only a fraction of the entire 3-M data base, which now contains on the order of 10^{11} bits (10 years' complete

Please enter your question.....

>> What types of aircraft are there.

parsing.....

THE FOLLOWING PLANES WILL BE RECOGNIZED BY THE SEMANTIC NET:

- (1) A3
- (2) A7
- (3) F4
- (4) F11
- (5) CHEROKEE
- (6) PHANTOM
- (7) SKYHAWK

The above planes may be further specified by giving the "tail number" (i.e. BUSER, BUNO, Bureau Serial Number, etc.) along with the name of the series.

{cpu time was 2.59 seconds, real time was 5.75 seconds.

Figure 5. PLANES' ability to answer questions about its contents.

data on all U.S. Navy aircraft, plus summaries).

Helpful Factors in the PLANES World

A number of factors contribute to making our problem much easier to solve than the general problem of understanding unconstrained natural language.

(1) Lack of ambiguity - Relatively few words and virtually no sentences in the PLANES world are ambiguous. The only ambiguous words we have been able to find are "wing" (meaning "a squadron" or "part of a plane") and "flight" (meaning "a flying event" or an adjective, as in "flight computer" or "flight director"). This means that if PLANES can find any interpretation at all for a request, it is in all likelihood to correct interpretation.

(2) Small vocabulary - Our current system has about 1100 words. We estimate that 2000 words will cover 90% or more of all requests made by users with at least a little prior experience with PLANES.

(3) Only two modes - PLANES is always either answering a question from the data base or attempting to help a user express his request in a form PLANES can understand. In general the system need not deal with declarative sentences.

(4) People do not type complex sentences - The increasing likelihood of making typing errors in lengthy requests, the increasing likelihood that long requests will baffle a program in some aspect, and general laziness all contribute to keeping input requests short and simple in construction. Reference 3 describes an experiment in which non-programmers thought that they were communicating with an intelligent program, when in fact they were interacting with another person who would respond appropriately to any input. It was found that 10 simple sentence types covered 78% of all input requests, and that another 10 would handle all but 10% of the requests.

(5) Less than 100% answer rate is acceptable - We feel that a 90% answer rate without rephrasing would be adequate to keep a user's interest and provide a practical and useful system. It is possible that even a lower rate might be acceptable.

(6) We have a good idea of what potential users would like to know - The Navy has made a study of all the requests made to 3-M data base during a one month period, and the frequency with which various requests were made. We thus had a good idea about where to concentrate our initial efforts, and the order in which to proceed.

Not-So-Helpful Factors in The PLANES World

(1) The System must contain a great deal of specialized knowledge -

one of our major realizations has been that a small number of general rules cannot suffice to "translate" natural English requests into data base queries. Consider the sentences:

(S1) "Which A7 has the worst maintenance record?"

(S2) "Find any common factors of plane numbers 37 and 78."

Clearly the system must contain special programs to compile a "maintenance record", and special knowledge to judge its "goodness"; the system must know that having the same digit as the fourth element of their serial numbers does not constitute a "common factor", but that similar event sequences are important.

(2) Each request may be expressed in a great many different ways. Clearly, if users are encouraged to sit at a console with little or no prior training or instruction, and if the system is expected to understand enough of a user's input to keep his interest and perform useful actions from the beginning, then the system must be able to make some sense of a large number of types of queries, and a wide range of syntactic constructions.

CONCLUSION

This paper has presented only a brief discussion of some design issues and examples of the performance of a natural language front end for a large data base. Many issues remain, both in the man/machine interface and in the underlying understanding mechanisms, but it is clear that we can now think of the design of such systems in engineering terms - such systems are no longer in the realm of science fiction.

Readers interested in more details of implementation and operation can consult reference 3; a technical report describing the system much more completely will be available in August 1976.

REFERENCES

1. Waltz, D. L. "Natural Language Access to a Large Data Base," Naval Research Reviews, Vol. XXIX, No. 1, Jan. 1976, pp. 11-25. (Also reprinted in Computers and People, Vol. 25, No. 4, April 1976.)
2. Codd, E. F. "Seven Steps to Rendezvous with the Casual User," Proc. IFIP TC-2 Working Conf. on Data Base Management Systems, Cargese, Corsica, April 1974, North-Holland Publ. Co., Amsterdam.
3. Malhotra, A. "Knowledge-Based English Language Systems for Management Support: An Analysis of Requirements," Advance Papers of the 4th Intl. Joint Conf. on Artificial Intelligence, Tbilisi, USSR, 1976, pp. 842-7.

EFFECTS OF LINGUISTIC REDUNDANCY ON PILOT'S

COMPREHENSION OF SYNTHESIZED SPEECH

Carol A. Simpson*

Ames Research Center, NASA, Moffett Field, California

SUMMARY

Previous research indicates that pilots' familiarity with aircraft phraseology results in high intelligibility for synthesized speech cockpit messages presented as short sentences - 95 percent articulation (percent correct) for unknown messages heard once (ref. 1). In the present study, the contribution of the sentence context to intelligibility was tested using two levels of linguistic redundancy: (1) two key words in isolation, for example, "*cargo fire*," versus (2) the same two key words in a short sentence context, for example, "You have *cargo* compartment *fire*."

In the first phase, airline pilots listened to unfamiliar messages that differed in linguistic redundancy, and that were presented in a background of continuous weather broadcast at seven signal-to-noise ratios (S/N) ranging from -8 dB to +8 dB. Pilot performance was assessed by the accuracy of comprehension for the two key words in each message and the response times measured from the end of each message to the pilot's signal of comprehension. In the second phase, pilots first studied a list of the same messages while listening to them and were then given recognition tests with and without competing weather broadcast at S/N of 0 dB.

Analysis of the data showed that monosyllabic key words in two-word format were significantly harder to understand than the same key words heard in short sentences. Key words in sentences were approximately 20 percent more intelligible than key words presented alone and response times to sentences were approximately 1 sec shorter than response times to two-word messages across all S/N conditions. For polysyllabic key words, the differences in intelligibility and response times for the two levels of linguistic redundancy were smaller but in the same direction.

It was concluded that synthesized speech messages for the cockpit should be worded in the form of short sentences or phrases, such as "The *fuel* pressure is *low*," instead of in abbreviated key word format, such as "*fuel low*." This design principle is especially important when the key words are monosyllabic.

*National Research Council Associate

INTRODUCTION

Miller, Heise, and Lichten (ref. 2) presented convincing evidence that words in a sentence context are more intelligible than the same words in isolation. Their finding applies to speech heard without the benefit of any particular real-world context. The authors also found that as the number of possible alternative interpretations decreased, due to the context of the situation or to knowledge of the message set, message intelligibility increased. These findings could be extended to predict that listeners familiar with a particular situation, such as flying an aircraft, and familiar with a particular type of well-learned phraseology, such as Air Traffic Control (ATC) phraseology and cockpit terminology, would exhibit, for messages pertaining to the flight situation, little or no intelligibility differences between words presented in isolation and words presented in sentences. It has been found in a previous study (ref. 1) that pilots do indeed have an advantage over nonpilots in understanding flight phraseology. The messages in that study were short ATC clearances and cockpit warnings and advisories. For both human and synthesized speech, pilots familiar with the type of phraseology, but not the particular message set, understood the cockpit messages better than did a control group of nonpilot police patrolmen, even though both groups were equally able to understand messages relating to commonly experienced events for both human and synthesized speech.

The wording of cockpit voice communications - including voice warnings - must be chosen such that the messages are understood accurately, rapidly, with few repetitions, and with as little demand on pilot mental workload and attention as possible. The present study was designed to determine whether the influence of familiarity with flight phraseology would be so strong as to outweigh the advantage that the sentence context was found to have in the Miller, Heise, and Lichten study (ref. 2) where the listeners were airline pilots presented with synthesized speech cockpit warnings. If short sentences, with their additional linguistic redundancy, fulfilled the requirements listed above better than did the key-word-only messages, then the longer, sentence-length format would be preferable for cockpit voice warnings.

A solution to this question is crucial for the following reasons. Brevity has been highly valued for speech communications in the cockpit because it is believed to reduce the length of the comprehension period. It does reduce the probability of temporal overlap with other auditory events. Also, brevity of wording conforms to the style of current cockpit communications with ATC and inside the cockpit for checklists and crew conversations. Finally, for warnings presented in digitized speech, each additional word represents additional cost in terms of dollars and additional read-only-memory. A corresponding additional hardware and memory cost for extra words for synthesized speech is not a problem since additional words are obtained via software modifications and memory increments on the order of only 100 bits per spoken word. The human factors "cost" for using brevity in cockpit voice warnings may be low intelligibility and high demands on pilot attention for comprehension.

EXPERIMENTAL DESIGN

For a comparison of the effectiveness of key-word-only messages to sentence-length messages, airline pilots listened to both types of message wording with and without competing weather broadcast. The study was divided into two phases. For Phase 1, the pilots were given no familiarization with the message set before testing and received no feedback on successive trials. For Phase 2, using the same messages, pilots first obtained familiarization with the warnings by listening to them while reading them from a printed list. Then the pilots were tested for their ability to recognize the warnings with and without competing weather broadcast.

Subjects

Sixteen airline pilots participated in Phase 1. Seven of them returned 6 months later for Phase 2, along with an eighth pilot who had participated in earlier voice warning research using the same speech synthesizer. The pilots' ages ranged from 31 to 44 years. They were paid for their participation.

Test Materials

Sixteen two-word warnings relating to airline transport operations were composed. Eight messages contained monosyllabic words and eight contained polysyllabic words. For each two-word warning, an appropriate sentence-length message was constructed that contained the same two key words presented in the same order as in the two-word version. Both versions of the messages are presented in table 1.

TABLE 1.- COCKPIT WARNING MESSAGES AND MESSAGE DURATIONS IN SECONDS.

MONOSYLLABIC WORDS	DURATION	MONOSYLLABIC WORDS IN SENTENCES	DURATION
1 FUEL LOW	0.55	1. THE FUEL PRESSURE IS LOW	1.68
2 GEAR DOWN	52	2. THE LANDING GEAR IS DOWN	1.75
3 CHECK FLAPS	66	3. CHECK YOUR FLAPS SETTING	1.90
4 BRAKES FAILED	65	4. THE WHEEL BRAKES HAVE FAILED	1.21
5 MAIN DOOR	59	5. THE MAIN DOOR IS NOT LOCKED	1.28
6 BOOST PUMP	54	6. THE FUEL BOOST PUMP IS OUT	1.39
7 OIL HOT	57	7. THE OIL TEMPERATURE IS HOT	1.59
8 ICE VALVES	47	8. THE WING ANTI ICE VALVES ARE INOPERATIVE	2.19
MEAN	57	MEAN	1.62

POLYSYLLABIC WORDS	DURATION	POLYSYLLABIC WORDS IN SENTENCES	DURATION
1. AUTOPILOT DISENGAGED	0.98	1. THE AUTOPILOT IS DISENGAGED.	1.36
2. DME MALFUNCTIONING	1.52	2. YOUR DME IS MALFUNCTIONING.	2.15
3. CABIN PRESSURE	61	3. THE CABIN PRESSURE IS LOW	1.68
4. SPOILERS INOPERATIVE	1.18	4. YOUR SPOILERS ARE INOPERATIVE	2.23
5. TERRAIN CLOSURE	86	5. YOU HAVE RAPID TERRAIN CLOSURE	2.18
6. CARGO FIRE	74	6. YOU HAVE CARGO COMPARTMENT FIRE	2.42
7. WINDOW OVERHEATED	65	7. THE COCKPIT WINDOW IS OVERHEATED	1.51
8. HYDRAULIC FAILURE	1.05	8. YOU HAVE HYDRAULIC SYSTEM FAILURE	1.89
MEAN	95	MEAN	1.93

The warnings were synthesized on a VOTRAX VS-6 speech synthesizer¹ using identical phoneme coding for the key words in their two-word and in their sentence-length versions. This coding ensured that the key words were pronounced identically in both types of message wording. A tape of the messages can be obtained from the author.

Continuous weather broadcast, used as background "noise", was recorded from a local weather station operated by the National Weather Service. This recording was then edited to eliminate pauses in the speech.

Testing Environment

Pilots were tested individually while sitting in a sound-treated booth. The warning messages and the weather broadcast were individually adjusted for desired peak sound pressure level (SPL)² and fed through an audio mixer. The summed output of the mixer was amplified and presented to the pilots binaurally via Koss ESP-9 electrostatic headphones. The pilots sat at a table which contained a response box with a button for signaling message comprehension (used in Phase 1) and a push-to-talk microphone for message readback (used in Phase 2). A PDP-12 computer was programmed to deliver the weather broadcast and the warnings and to record response times.

Experimental Conditions

Phase 1- For Phase 1, the pilots had no prior experience with the message set. Each pilot was assigned a message set containing eight two-word warnings and eight sentence-length warnings. Pilots were alternately assigned to one of two groups. Group 1 pilots were given two-word *monosyllabic* warnings and sentence-length warnings containing *polysyllabic* words. Group 2 pilots heard two-word *polysyllabic* warnings and sentence-length warnings containing *monosyllabic* key words. In this way, each pilot was tested on all 16 pairs of key words but no pilot received the same key-word pairs in both the two-word and the sentence-length versions. For a baseline comparison of the ability of the two groups to understand the speech produced by the synthesizer, all the pilots were given a pre-test consisting of 16 unfamiliar short ATC clearances and a post-test containing 16 unfamiliar cockpit advisories, without competing weather broadcast.

Each pilot was given repeated trials on his entire message set of eight two-word and eight sentence-length warnings at seven levels of increasing signal-to-weather broadcast (S/W) and peak SPL ratios of -8, -5, -3, 0, +3, +5, and +8 dB. The weather broadcast was held constant at 60 dB SPL, and the SPL of the warnings was varied from 52 dB to 68 dB to obtain the desired S/W. For the eighth and final trial, the 16 warning messages were presented at 60 dB SPL without weather broadcast. The experimental conditions for Phase 1 are shown in Table 2.

¹Invented by Richard T. Gagnon and manufactured by Vocal Interface Division of Federal Screw Works, Troy, Michigan.

²Referenced to 0.0002 dyne/cm².

TABLE 2.- EXPERIMENTAL CONDITIONS
FOR PHASE 1.

	8 TWO WORD WARNINGS	8 SENTENCE WARNINGS
GROUP 1 8 AIRLINE PILOTS	MONOSYLLABIC KEY WORDS	POLYSYLLABIC KEY WORDS IN SENTENCES
GROUP 2 8 AIRLINE PILOTS	POLYSYLLABIC KEY WORDS	MONOSYLLABIC KEY WORDS IN SENTENCES

LISTENING CONDITIONS									
WARNING SPL (dB)	52	55	57	60	63	65	68	60	
WEATHER SPL (dB)	60	60	60	60	60	60	60	-	
SN	-8	-5	-3	0	3	5	8	40	
TRIAL No	1	2	3	4	5	6	7	8	

SOUND PRESSURE LEVELS IN DECIBELS RE 0.002 dyne/cm²

The order of presentation of the two-word messages and the sentence-length messages was balanced across pilots so that half the pilots in each group heard sentences before the two-word format for all trials and the remainder heard two-word format before sentences. The messages were presented to each pilot in a different random order for each trial. As they arrived for testing, pilots were assigned alternately to either the sentences-first order or to the words-first order.

For trials 1 through 7, continuous weather broadcast occurred throughout each trial. For each SPL, warning messages were presented one at a time at 40-sec intervals. Pilots responded by pressing a button labeled "Understand" as soon as they believed they understood

each message. They then wrote the message on a prepared answer sheet. If they did not understand a message, they were instructed to think about it for a while and then press the button and either write their best guess or the word "no." Response times from the end of each message to the pilot's "Understand" button depression were recorded. The procedure for the eighth and final trial and for the pre- and post-tests was the same except that no competing weather broadcast was used and the SPL for all messages was 60 dB.

Phase 2- The same messages used in Phase 1 were also used in Phase 2. Four pilots were assigned to the monosyllabic two-word messages and the polysyllabic sentence-length messages. The other four pilots heard the polysyllabic two-word messages and the monosyllabic sentence-length messages. Assignment of pilots to message sets was such that the seven pilots originally in Phase 1 heard the opposite message set in Phase 2 than they had heard in Phase 1. As in Phase 1, the order of words-first versus sentences-first was balanced across pilots. Each pilot was given a different presentation order in each of three listening conditions. The first listening condition was a familiarization session. While looking at a printed message list, the pilot heard each message presented twice in succession. Next, he received a recognition test in which the messages were presented only once at 12-sec intervals in a different order and with no competing weather broadcast. The pilot responded by keying a microphone and reading back each message as rapidly as possible, then writing it down on an answer sheet. Finally, he was tested on the same messages presented in a third order and embedded in weather broadcast at a S/W of 0 dB. Again, the pilot responded by keying a microphone and then reading back and writing down the message. The weather broadcast and warning message were delivered concurrently with the weather leading the onset of the warning by approximately 950 msec and continuing until the end of the warning. An interval of 40 sec was used to reduce the changes that the pilots would be prepared for successive presentations. All messages and the weather broadcast

TABLE 3.- EXPERIMENTAL CONDITIONS FOR PHASE 2.

8 TWO WORD WARNINGS		8 SENTENCE WARNINGS
4 AIRLINE PILOTS	MONOSYLLABIC WORDS	POLYSYLLABIC WORDS IN SENTENCES
4 AIRLINE PILOTS	POLYSYLLABIC WORDS	MONOSYLLABIC WORDS IN SENTENCES

LISTENING CONDITIONS			
	LISTEN TO WARNINGS WHILE READING LIST	RECOGNIZE WARNINGS WITHOUT WEATHER	RECOGNIZE WARNINGS DURING WEATHER S N
WARNING SPL dB	50	50	50
WEATHER SPL dB	-	-	50
TRIAL No	1	2	3

were delivered at approximately 50 dB SPL. Table 3 shows the experimental conditions for Phase 2.

RESULTS

Group Differences

Wilcoxon's Sum of Ranks for the Pre-test ($R = 68$, $n_1 = 8$, $n_2 = 8$, $p > 0.10$) and the Post-test ($R = 68$, $n_1 = 8$, $n_2 = 8$, $p > 0.10$) indicated that the two groups of pilots were not significantly different in their ability to understand the synthesizer. Means and standard deviations for the two groups for pre- and post-test articulation scores (percent correct) are shown in Figure 1. However, the two groups were significantly different ($t = 2.27$, $df = 14$, $p < 0.05$) in average response times to the pre-test messages. They were not significantly different ($t = 1.16$, $df = 14$, $p > 0.10$) in average response times for the post-test messages. Table 4 shows the means and standard deviations in response time for the two groups for the pre- and post-test. Because of the initial response time difference between the two groups, which may have continued for an indeterminate number of trials on the test messages, the response time data was normalized using individual subjects' means and standard deviations across message type so that between-group comparisons could be made for the same key-word pairs presented in isolation and in sentence context.

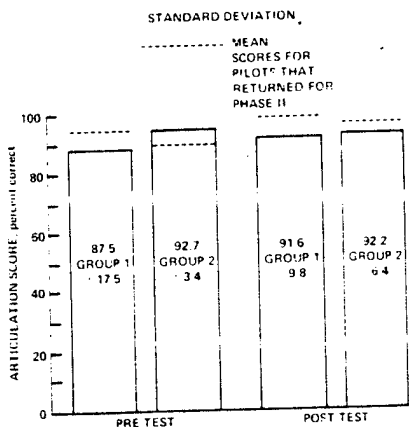


Figure 1.- Mean articulation scores (percent correct) for the two Pilot Groups for synthesized speech Pre- and Post-tests. N = 8 pilots per group.

TABLE 4.- COMPARISON OF RESPONSE TIMES FOR PRE- AND POST-TEST MESSAGES FOR PILOT GROUPS 1 AND 2.

	PRE TEST	POST TEST
GROUP 1 8 AIRLINE PILOTS	3.51 sec 2.37 sec	1.67 sec 0.80 sec
GROUP 2 8 AIRLINE PILOTS	1.66 sec 0.58 sec	1.31 sec 0.55 sec

Presentation Order

Analysis using Wilcoxon's Sum of Ranks determined that presentation of sentences or words first had no significant effect on the articulation scores for words ($R = 59.5$, $n_1 = 8$, $n_2 = 8$, $p > 0.10$) or for sentences ($R = 68$, $n_1 = 8$, $n_2 = 8$, $p > 0.10$). Similarly, there were no significant presentation

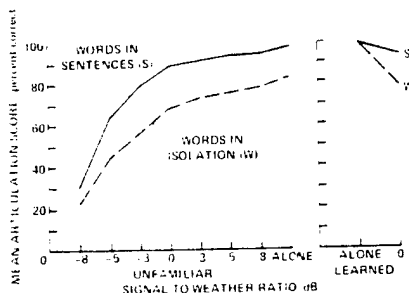


Figure 2.- Mean articulation scores for key monosyllabic words in unfamiliar synthesized speech cockpit warnings with and without a sentence context, heard in background of continuous weather broadcast for repeated trials at increasing signal-to-"noise" ratios (Phase 1) and for the same messages with and without competing weather broadcast after prior learning of the message set (Phase 2).

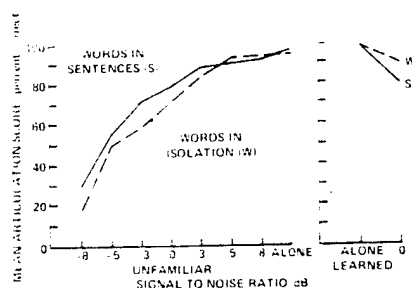


Figure 3.- Mean articulation scores for key polysyllabic words in unfamiliar synthesized speech cockpit warnings with and without a sentence context, heard in background of continuous weather broadcast for repeated trials at increasing signal-to-"noise" ratios (Phase 1) and for the same messages with and without competing weather broadcast after prior learning of the message set (Phase 2).

order effects on response times to words ($R = 52$, $n_1 = 8$, $n_2 = 8$, $p > 0.10$) or for response times to sentences ($R = 58$, $n_1 = 8$, $n_2 = 8$, $p > 0.10$).

Articulation Scores

Articulation scores are the percent of key words correctly identified for each message wording type.

Monosyllabic words- Figure 2 shows a between-groups comparison of articulation scores for monosyllabic key words in the two-word and sentence-length versions. The eight trials at increasing S/W for Phase 1 are on the right, and the two testing trials of Phase 2 are on the left. The higher mean articulation scores indicated that for unfamiliar messages, the key words in a sentence context were significantly more intelligible than the same key words presented alone across all S/W conditions, significance being determined by Wilcoxon's Signed Ranks Test ($R = 0$, no. of pairs (n_p) = 8, $p < 0.01$). In Phase 2, when pilots were given prior familiarization with the message set and tested for recognition of the "learned" messages, all pilots scored 100 percent articulation for all messages of both types of wording when no competing weather broadcast was present. With weather broadcast present, mean articulation for words presented alone dropped to 78 percent, but only dropped to 94 percent when the same words were presented in a sentence context. This difference, however, did not reach significance as measured by Wilcoxon's Signed Ranks Test ($R = 2.5$, $n_p = 6$, $p > 0.10$). Only six pairs were used because the difference scores within two of the pairs were zero.

Polysyllabic words- The mean articulation scores for unfamiliar polysyllabic words with and without sentence context in Phase 1 are similar to those for the monosyllabic words at S/W of +3 dB and lower, as shown on the left side of figure 3. Across all S/W conditions, however, the

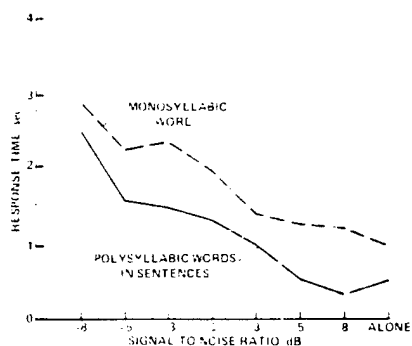


Figure 4.- Response times for Pilot Group 1 for cockpit warnings: Monosyllabic words alone and Polysyllabic words in sentences presented for repeated trials at increasing signal-to-weather ratios. Response times measured from end of warning to pilot's "Understand" signal (Phase 1).

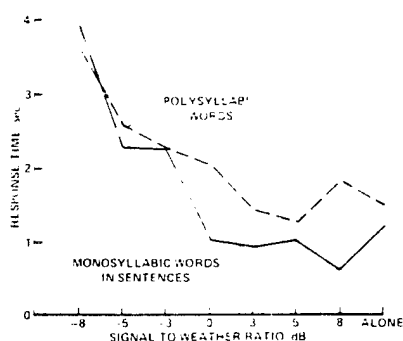


Figure 5.- Response times for Pilot Group 2 for cockpit warnings: Polysyllabic words alone and Monosyllabic words in sentences presented for repeated trials at increasing signal-to-weather ratios. Response times measured from end of warning to pilot's "Understand" signal (Phase 1).

addition of a sentence context did not significantly improve ($R = 3$, $n_p = 8$, $p > 0.10$) intelligibility for polysyllabic key words in initially unfamiliar messages. For previously learned warnings (Phase 2), articulation scores were again 100 percent for both types of wording presented with no competing weather broadcast. When weather broadcast was presented at S/W of 0 dB, articulation scores for words presented alone dropped to 92 percent and articulation for words in sentences dropped to 81 percent. The difference between articulation scores for words alone and in a sentence context was again not significant ($R = 4.5$, $n_p = 6$, $p > 0.10$).

Response Times

The response-time data closely paralleled the intelligibility data in Phase 1. The response times were measured from the end of each message to the time when the pilot pressed the "Understand" button and began his response, regardless of accuracy. For Pilot Group 1 (fig. 4) and Pilot Group 2 (fig. 5), the mean response time for the sentence-length messages was significantly shorter (Group 1: $R = 0$, $n_p = 7$, $p < 0.05$; Group 2: $R = 3$, $n_p = 8$, $p < 0.05$) than the mean response time for the two-word messages. For Group 1, response time data was available for only 7 of the 8 pilots since one pilot misunderstood the instructions and pressed the "Understand" button after he finished his responses.

Included in the response time data are several types of responses:

- 1) incorrect response,
- 2) "unable" response,
- 3) message correct for the first time,
- 4) message correct on a previous trial and correct again on the current trial.

It would be reasonable to expect differences in response times as a function of the type of response the pilot made. Therefore, mean

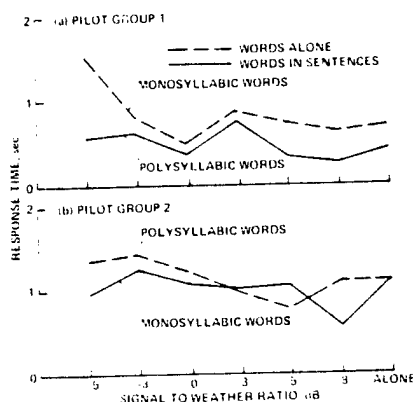


Figure 6.- Response times for cockpit warnings recognized correctly and recognized correctly on previous trial for repeated trials at increasing signal-to-weather ratios (Phase 1).

response times were calculated only for responses of type 4, that is, for messages which a given pilot had understood correctly and had also understood correctly on the previous trial at a lower S/W. The assumption was that for the small message set used in this study, once a pilot correctly understood a particular warning, he was likely to remember it during the next trial and so facilitate the recognition process. The response time to such "previously understood and now familiar" messages should allow a valid comparison between words presented in isolation and those presented in a sentence context, for familiar messages. In general, the response times to the familiar, correctly recognized warnings were shorter for both Pilot Groups (figs. 6(a) and 6(b)) than were the response times to all messages regardless of accuracy (figs. 4 and 5).

Because of group differences in mean response times, each pilot's response times were normalized. The resulting z-scores were then used for between-group comparisons of the effects of sentence context on monosyllabic (fig. 7) and polysyllabic (fig. 8) words. In each case, the response times for each S/W were significantly longer (Monosyllabic words: $R = 0$, $n_p = 8$, $p < 0.01$; Polysyllabic words: $R = 0$, $n_p = 8$, $p < 0.01$) for the two-word messages than for the sentence-length messages. The size of the response time difference diminished slightly for the polysyllabic words at positive S/W conditions, but the difference occurred at all S/W conditions for both mono- and polysyllabic words.

Four-way Comparison of Message Types

In addition to the major comparisons of articulation scores and response times for the mono- and polysyllabic words with and without a sentence context,

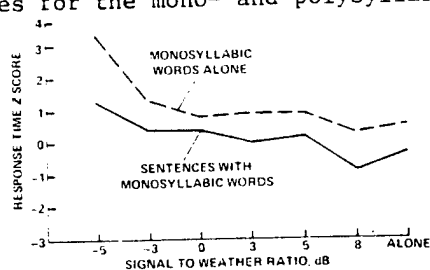


Figure 7.- Z-scores for response times for "familiar" cockpit warnings containing monosyllabic words, with and without sentence context, presented for repeated trials at increasing signal-to-weather ratios (Phase 1).

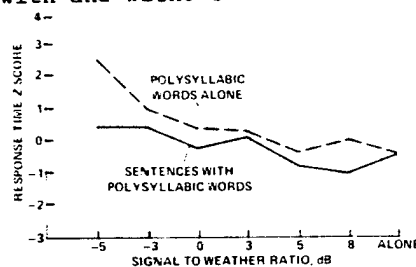


Figure 8.- Z-scores for response times for "familiar" cockpit warnings containing polysyllabic words, with and without sentence context, presented for repeated trials at increasing signal-to-weather ratios (Phase 1).

it is interesting to note that pilot performance for monosyllabic words presented alone (MW) is poorer than performance for the other three presentation types: monosyllabic words in sentences (MS), polysyllabic words alone (PW), and polysyllabic words in sentences (PS). Peak intelligibility for MW was only 84 percent compared to 98 percent for MS (fig. 2) and 95 percent and 97 percent for PW and PS, respectively (fig. 3). Similarly, the z-scores for response times to familiar messages for the last trial are well above the mean for MW (+0.46) and below the mean for MS (-0.37) (fig. 7), PW (-0.56), and PS (-0.58) (fig. 8). These data indicate that response times to the monosyllabic words presented alone were consistently longer than response times to polysyllabic words presented alone and both mono- and polysyllabic words presented in sentences.

DISCUSSION

The reader is cautioned against an interpretation of intelligibility and response time as a direct function of S/W, because this function is confounded with learning. The experimental paradigm used in this study makes use of the modified method of limits, also known as the ascending method of limits (ref. 3). It is recommended for obtaining only the *relative* thresholds for some variable, such as S/N ratio, for two or more classes of stimuli which are to be identified by the subject without prior knowledge of the total set of response alternatives. By presenting repeated trials on alternating lists of the different classes of stimuli at increasing intensities, durations, or S/N ratios, the relative thresholds for the different classes of stimuli under equal conditions of number of previous trials can be obtained while minimizing the effects of subject familiarity with the individual stimuli. If anything, the effects of increasing familiarity with the message set would be expected at higher levels of S/W to reduce the chances of significant intelligibility and response-time differences between the two levels of linguistic redundancy for message wording.

Nevertheless, intelligibility scores are higher and reaction times are shorter for synthesized-speech cockpit warnings presented as short sentences than they are for warnings presented in a briefer two key-word format. For listeners who are experienced with cockpit phraseology, this relationship is true across a wide range of S/W broadcast ratios. The advantage of the sentence context is larger for monosyllabic words than for polysyllabic words in that for polysyllabic words the facilitating effect of the added linguistic redundancy decreases or perhaps disappears at S/W of +3 dB or higher. Familiarity with individual messages decreases reaction time and increases intelligibility but does not override the facilitating effect of the sentence context for recognition and response times. Monosyllabic words alone produce consistently poorer performance in terms of lower intelligibility and longer response times than do polysyllabic words alone and both monosyllabic and polysyllabic words in sentences. This performance relationship is consistent with Howe's research (ref. 4), in which longer words were found to be more intelligible than shorter words, presumably because of their phonetic redundancy - there are far more restrictions on allowable sequences of sounds that "make sense" when additional syllables are included.

It would seem, then, that the additional linguistic redundancy provided by the sentence context, and to a lesser degree by the additional syllables in the polysyllabic words, helps pilots process synthesized warning messages more accurately and more quickly than when the two-word format is used. This conclusion is further supported by a subsequent study, also presented at the 12th Annual Conference on Manual Control (ref. 5) and using time estimation as a secondary task, in which it was found that the more redundant sentences required less attention for processing. It was also found that *total* listening time - measured from the beginning of the warning with the warnings repeated until the pilot cancelled the message to signal that he understood - was shorter for sentences than for two-word messages, because the pilots allowed the two-word warnings to repeat more times than the sentence-length warnings. This difference occurred despite a ratio of sentence to two-word warning-message duration 2.3 to 1. The operational implications of this finding are strengthened by the fact that the actual duration of any sentence warnings synthesized for use in the cockpit would be slightly shorter than the durations used in this study. The durations of the key words in this study were made identical for the words-alone and the words-in-sentences versions of the warnings so as to present identical acoustic stimuli for both conditions of linguistic redundancy. In human speech, the pronunciation of all but the final words in sentences and phrases is actually shorter than when the words are pronounced in isolation (ref. 6). Additionally, the shortening of the key words in the sentences would improve timing rhythm of the sentence pronunciation, which would be expected to result in still higher sentence intelligibility because of the perceptual importance of proper timing relationships among the segments of sentences (ref. 7).

If the increased recognition accuracy and decreased response time for the sentences versus the two-word warnings is taken as an indicator of less cognitive effort for processing the sentences, then it is possible that under extremely high workload conditions in the cockpit, sentence messages would be understood more reliably by conveying the warning more accurately and rapidly; possibly with only a single repetition.

This suggestion is supported by judgments from the eight pilots who participated in Phase 2. They were asked to judge the difficulty of learning and recognizing the two types of message wording in the experimental situation and to give their preferences for warning-message wording in the cockpit. While most of the pilots found both types of wording equally easy to learn and to recognize in the absence of competing speech, five of the eight thought that the sentences were easier to recognize during simultaneous weather broadcast, and six of the eight would prefer the sentence-type wording over the two-word format for cockpit warnings (table 5). Their reasons for this preference were obtained in response to an open-ended question, "Why do you think the _____ were easier than the _____?" The reasons the pilots gave for preferring sentences to two-word messages fell into three general categories:

- 1) Extra words in the sentence help you figure out the words you missed.
- 2) The first words of the sentence get your attention and cue you in to the particular voice that is talking so that you are listening for the critical words.

TABLE 5.- AIRLINE PILOT PREFERENCES FOR COCKPIT WARNING MESSAGE WORDING: TWO KEY WORDS VS A SENTENCE (PHASE 2).

SITUATION		TWO WORD FORMAT EASIER	SENTENCE FORMAT EASIER	NO DIFFERENCE	N
	MESSAGE LEARNING	2	0	6	8
	RECOGNITION WITH NO COMPETING WX	1	2	5	8
	RECOGNITION DURING COMPETING WX	0	5	3	8
	PREFERENCE IN COCKPIT	2*	6	0	8

*THE TWO PILOTS WHO PREFERRED THE TWO WORD FORMAT FOR USE IN THE COCKPIT THOUGHT THE LONGER MESSAGES WOULD DISTRACT THEM FROM OTHER FLIGHT TASKS. BOTH FOUND THE FORMAT WARNING • KEY WORD • KEY WORD

- 3) The longer pattern of the sentence with extra words between the critical ones gives you more time to understand the words.

The two pilots who preferred two-word warnings expressed a preference for brevity of verbal communications. One thought that the sentences, being longer, would distract his attention more from his flight tasks (but see ref. 5). Both these pilots, however, thought that a warning message format would be acceptable if it consisted of an alerting word such as "warning" followed by a sequence of key words.

TABLE 6.- AIRLINE PILOT RESPONSES TO A PROPOSED COCKPIT VOICE-WARNING SYSTEM (PHASE 2).

REPLACE ALL CURRENT AURAL WARNINGS WITH VOICE
AUTOMATIC VOLUME CONTROL (AVC) TO MAINTAIN CONSTANT
S.N RATIO

PRIMARY DELIVERY VIA HEADSETS AT LEVEL EQUAL TO
INDIVIDUAL PILOT SELECTED COMMUNICATIONS AUDIO LEVEL

BACK UP DELIVERY VIA SPEAKER USING AVC TO MAINTAIN
CONSTANT SIGNAL TO COCKPIT NOISE LEVEL

PILOT PREFERENCE

YES		NO		N	
ELIMINATE ALL BELLS, HORNS, CLACKERS, BUZZERS	KEEP FIRE BELL ONLY	VOICE OK FOR SOME WARNINGS	NO VOICE		
No. OF PILOTS	4	3	1	0	8

TABLE 7.- AIRLINE PILOT PREFERENCES FOR REPETITION OF COCKPIT VOICE WARNINGS (PHASE 2).

SHOULD VOICE WARNINGS BE REPEATED?

YES	NO	N
8	0	8

NUMBER OF REPEATS FOR COCKPIT VOICE WARNINGS

CRITICALITY OF WARNING SITUATION	1	2 OR 3	UNTIL CANCELLED OR PROBLEM CORRECTED	UNTIL PROBLEM CORRECTED - NOT CANCELLABLE	N	
				(1) LANDING GEAR AND FLAPS ONLY		
	CRITICAL	0	1	7		8
NON CRITICAL	0	4	4	0		8

Table 7 deals with warning repetition and cancellation. All eight pilots thought that the warnings should be repeated more than once. Preferences for the number of repeats and type of cancellation depended to some extent on the criticality of the warnings. For critical warnings, seven of the

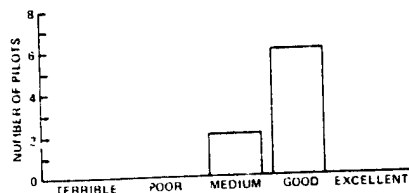
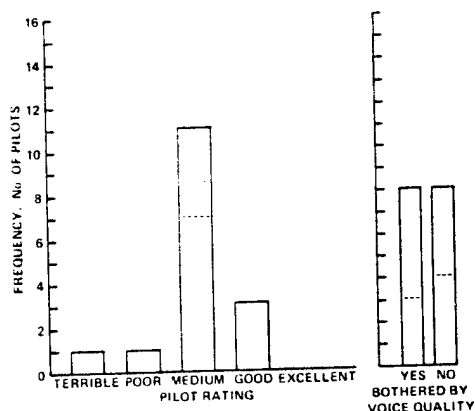


Figure 9.- Airline pilot ratings of clarity of pronunciation of synthesized-speech cockpit warnings after learning message set and recognizing warnings with and without competing weather broadcast (Phase 2).

TABLE 8.- AIRLINE PILOT ACCEPTABILITY OF SYNTHESIZED SPEECH AND DIGITIZED SPEECH FOR COCKPIT VOICE WARNINGS (PHASE 2).

	YES	NO	N
WOULD WANT THIS VOICE IN OWN COCKPIT	7	1	8
WOULD WANT GPWS TYPE VOICE IN OWN COCKPIT	8	0	8



DOTTED LINES INDICATE PREFERENCES OF THE 7 PILOTS WHO RETURNED SIX MONTHS LATER FOR TESTING ON THE SAME MESSAGES WITH PRIOR FAMILIARIZATION WITH MESSAGE SET

Figure 10.- Airline pilot ratings of pronunciation clarity of initially unfamiliar synthesized-speech cockpit warnings heard at increasing signal-to-weather broadcast sound pressure level ratios (Phase 1).

eight pilots thought that the warnings should be repeated until either a crew member cancelled them or the situation was corrected, whichever happened first. One of these seven made an exception to this for the landing gear and flaps warning, stating that at low altitudes above the ground, he would not want to be able to cancel this warning until the problem was corrected.

For noncritical warnings, the pilots were divided in opinion with four preferring two or three repeats with automatic cut-off and the other four preferring to have the warnings repeated until cancelled either by a crew member or by correction of the problem.

These pilots judged the clarity of pronunciation of the messages produced on this synthesizer as generally good, and seven of the eight said that they would find this particular voice acceptable in their cockpit (fig. 9 and table 8). All eight pilots judged the voice quality of the Ground Proximity Warning System (GPWS) - representing digitized speech - acceptable but thought that the current GPWS voice is too loud. It is interesting to compare the responses of these pilots to their responses after the earlier testing in the unfamiliar message condition of Phase 1. Figure 10 shows the voice clarity ratings of the 16 pilots in the Unfamiliar Message Set condition. As a group, they tended to rate this synthesizer as having medium clarity of pronunciation, and half the group said they were "bothered by the voice quality." The dotted lines show the responses of the seven pilots who returned six months later for retesting on the same messages under the familiar condition. They were representative of the larger group, as evidenced by their medium ranking of pronunciation and divided opinion regarding voice quality. Several factors may have combined to cause them to change their rating of the pronunciation and to be willing to have this particular voice

in their own cockpit, such as familiarity with the particular messages and experience listening to the synthesizer voice.

CONCLUDING REMARKS

Advantages seemingly may be gained from carefully designed cockpit voice warnings. One design principle that should be followed in order to ensure the effectiveness of this mode of presenting warning information to pilots is to include sufficient linguistic redundancy in the message wording to ensure rapid and accurate comprehension. The context provided by a short sentence has been found to improve accuracy and decrease response times to cockpit warnings, particularly when the key words in the warning are monosyllabic. This effect is significant despite a strong facilitating effect of pilots' familiarity with flight phraseology. Further research will be directed at assessing the facilitating effects of the different types of linguistic redundancy - semantic and syntactic - and the alerting and "accent"-cueing effect of an initial low information word provided by the sentence context. These effects will be studied under simulated flight conditions with varying levels of pilot attention workload.

REFERENCES

1. Simpson, C. A.: Occupational Experience with a Specific Phraseology: Group Differences in Intelligibility for Synthesized and Human Speech. Abstract in the Journal of the Acoustical Society of America, vol. 58, Suppl. 1, 1975, p. 57. Paper presented at the 90th Meeting of the Acoustical Society of America, San Francisco, Nov. 5, 1975 and available from the author.
2. Miller, G. A.; Heise, G. A.; and Lichten, W.: The Intelligibility of Speech as a Function of the Context of the Test Materials. Journal of Experimental Psychology, vol. 41, 1951, pp. 329-335.
3. Dember, W. N.: The Psychology of Perception. Holt, New York, 1960, pp. 60-65.
4. Howes, D. H.: On the Relation Between the Intelligibility and Frequency of Occurrence of English Words. Journal of the Acoustical Society of America, vol. 29, 1957, pp. 296-305.
5. Hart, S. G.; and Simpson, C. A.: Effects of Linguistic Redundancy on Synthesized Cockpit Warning Message Comprehension and Concurrent Time Estimation. Paper presented at the 12th Annual Conference on Manual Control, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, May 25-27, 1976. To be published in the Conference Proceedings as a NASA TMX (in press).
6. Klatt, D. H.: Vowel Lengthening is Syntactically Determined in a Connected Discourse. Journal of Phonetics, vol. 3, 1975, pp. 129-140.
7. Dooling, D. J.: Rhythm and Syntax in Sentence Perception. Journal of Verbal Learning and Verbal Behavior, vol. 13, 1974, pp. 255-264.

EFFECTS OF LINGUISTIC REDUNDANCY ON SYNTHESIZED COCKPIT WARNING
MESSAGE COMPREHENSION AND CONCURRENT TIME ESTIMATION

Sandra G. Hart*
University of California, Berkeley, California

Carol A. Simpson**
Ames Research Center, NASA, Moffett Field, California

SUMMARY

The relative amounts of attention required to comprehend and recognize two types of speech messages that differed in linguistic redundancy were evaluated with two concurrent time estimation tasks. Two-word and sentence-length synthesized-speech cockpit warning messages were presented to 12 commercial airline pilots with and without competing weather broadcast. The pilots had two tasks: a) to read back and write down the warning messages and b) to give two types of time estimates. It was found that the intelligibility of messages presented in a redundant sentence-length format was higher, and that listening time and number of repetitions was less, during the Familiarization Phase. Sentences were also more intelligible than two-word messages when presented in a background of competing weather broadcast for recognition. Differences in verbal estimates of session length and in the length of 10-sec estimates produced by pilots indicated that messages presented in a sentence format required less attention for comprehension than did two-word messages.

INTRODUCTION

It might be expected that the many different concurrent tasks performed in flight would demand varying amounts of attention and mental workload, even though such variation might not be reflected in pilot performance on the primary flying task. Two types of time estimation tasks have been used as quantitative measures of such concurrent task processing requirements: a) verbal estimation in which the length of an interval of time that has been experienced is estimated in minutes or seconds and b) time production in

*This author's research was supported by NASA Grant NGL 05-046-002 to the San Jose State University Foundation.

**This author's research was supported by a post-doctoral fellowship from the National Research Council.

which the duration of an interval of time that is specified verbally is estimated by reproducing it operationally. Hart (ref. 1) and Hart and McPherson (ref. 2) reported that the length of time productions made during compensatory tracking and simulated flight, and the length of verbal estimates made of intervals filled with tracking, covaried with the difficulty of the manual control tasks. It was suggested that the length of verbal produced time estimates reflected the attention demands of concurrent activity. In the present study, similar time estimation tasks were used to assess the attention required for two speech processing tasks: a) comprehending unfamiliar two-word and sentence-length synthesized-speech cockpit warnings (Familiarization Phase), and b) recognizing these messages once they were familiar (Recognition Phase).

It has been previously reported that words presented in a sentence context are more intelligible than the same words in isolation, by virtue of the increased redundancy (refs. 3, 4). Linguistic redundancy refers to a reduction in the number of interpretations which can be assigned to the segments of an utterance due to semantic, syntactic, and phonological constraints such as those provided by a sentence frame. Since the results of Simpson's research (ref. 3) showed increased intelligibility and decreased response time as the redundancy in synthesized cockpit warnings was increased, it seemed reasonable to assume that the attention required of pilots to comprehend messages initially and recognize them once they were familiar would be inversely related to redundancy. The present study was designed to assess by means of verbal and produced time estimates, the attention required to comprehend and recognize warnings presented in either two-word or sentence formats. The operational motivation for this study was to examine the possibility that increased linguistic redundancy would reduce the attention required for pilots to process synthesized-speech warning messages in the cockpit. If warnings could be worded so as to reduce the attention required for their comprehension, this could leave more attention available for flying the airplane and taking proper corrective actions. This savings in attention would be especially important in emergency situations, where higher-than-normal attention demands exist.

It was predicted that less redundant two-word warnings would require longer listening time and more repetitions for initial comprehension and that they would be less intelligible than messages presented in a sentence format during both the Familiarization and Recognition Phases. If attention demands vary inversely with redundancy, as predicted, pilots should increasingly underestimate the passage of time as message redundancy is progressively decreased. (Underestimation occurs when concurrent activity momentarily diverts attention from timekeeping so clock time passes relatively unnoticed.) Increasing underestimation of time will in turn result in progressively shorter verbal estimates and longer produced durations.

TASK AND PROCEDURE

Subjects

Twelve commercial airline pilots, ranging in age from 35 to 45 years, served as paid participants in the present study.

Test Materials

The intelligibility of 16 monosyllabic and polysyllabic two-word synthesized-speech warning messages was assessed by computing the percentage of words correctly identified in a pilot study. Eight messages were then assigned to each of two message sets so that the average intelligibility of the sets was approximately equal. Sixteen sentence-length warning messages were then constructed from each of the sixteen two-word messages. As shown in table 1, each sentence-length warning contained the same two words in the same order as in its corresponding two-word version.

TABLE 1.— COCKPIT WARNING MESSAGES AND THEIR DURATION IN SECONDS

	GROUP 1 PILOTS		GROUP 2 PILOTS	
	TWO-WORD FORMAT	DURATION (Sec)	SENTENCE FORMAT	DURATION (Sec)
	FUEL LOW	.55	YOU HAVE HYDRAULIC SYSTEM FAILURE.	1.89
	CABIN PRESSURE	.61	THE FUEL BOOST PUMP IS OUT.	1.39
	GEAR DOWN	.52	THE COCKPIT WINDOW IS OVERHEATED.	1.51
	DME MALFUNCTIONING	1.52	THE WHEEL BRAKES HAVE FAILED.	1.21
	CHECK FLAPS	.66	THE AUTOPILOT IS DISENGAGED.	1.36
	SPOILERS INOPERATIVE	1.18	THE OIL TEMPERATURE IS HOT.	1.59
	TERRAIN CLOSURE	.86	THE WING ANTI-ICE VALVES ARE INOPERATIVE.	2.19
	CARGO FIRE	.74	THE MAIN DOOR IS NOT LOCKED.	1.28
	MEAN DURATION	.83	MEAN DURATION	1.55
	HYDRAULIC FAILURE	1.05	THE FUEL PRESSURE IS LOW.	1.13
	BOOST PUMPS	.54	THE CABIN PRESSURE IS LOW.	1.13
	WINDOW OVERHEATED	.65	THE LANDING GEAR IS DOWN.	1.20
	BRAKES FAILED	.68	YOUR DME IS MALFUNCTIONING.	1.60
	AUTOPILOT DISENGAGED	.98	CHECK YOUR FLAPS SETTING.	1.35
	OIL HOT	.57	YOUR SPOILERS ARE INOPERATIVE.	1.68
	ICE VALVES	.47	YOU HAVE RAPID TERRAIN CLOSURE.	1.63
	MAIN DOOR	.59	YOU HAVE CARGO COMPARTMENT FIRE.	1.87
	MEAN DURATION	.69	MEAN DURATION	1.45

The warnings were synthesized on a VOTRAX VS-6 speech synthesizer using identical phoneme coding for the key words in both the two-word and sentence-length message formats. A PDP-12 computer delivered the messages and recorded the response times.

The background noise used in the Recognition Phase was continuous weather broadcast recorded from a local station operated by the National Weather Service and edited to eliminate pauses. The ratio of warning-message signal peak sound pressure level (SPL) to weather-broadcast noise SPL was +3 dB.

Procedure

Each pilot was tested individually in a sound-attenuated booth. The warning messages and weather broadcast were presented to the pilots binaurally through Koss earphones, Model ESP-9. Pilots were asked to initiate message readback by keying a hand-held microphone. Subsequent to verbal readback, pilots were asked to write down on a prepared answer sheet as much as they had understood of the message. Message readback via keyed microphone was used to provide pilots with a response mode similar to that which they normally use for responding to spoken messages in the cockpit. It was also expected that vocalization would result in shorter response times than would be obtained using written responses. However, for intelligibility scoring, a written record of the pilot's response was required so as to eliminate possible experimenter bias in listening to recorded spoken responses.

Pilots made verbal time estimates during the Familiarization Phase by writing their estimate of elapsed time on the answer sheet. They produced 10-sec intervals during the Recognition Phase by pressing one button to begin their estimate when cued to do so by a pre-recorded message and a second button to terminate their production when they believed that 10 sec had elapsed.

Half of the pilots were randomly assigned to each of two test groups. The six pilots in one group received the first set of eight two-word messages and the eight sentence-length messages constructed from the second set of two-word warnings. Conversely, the second group of six pilots received the second set of two-word warnings and the sentence-length messages constructed from the first set of two-word warnings. The messages given to each group and their durations can be seen in table 1.

All twelve pilots were given three familiarization runs for each of eight two-word messages and each of eight sentence-length messages. After a 30-min break, pilots were given one recognition run in which each message was presented one time only in a background of continuous weather broadcast. The order of presentation of the two-word and sentence-length warning message sets was balanced across subjects so that three pilots in each group heard sentences before two-word messages during all the test runs. The other three pilots in each group heard the two-word versions first, and then the sentence-length versions. The eight two-word and eight sentence-length messages within the two sets were presented in a different random order for each pilot.

Familiarization phase— During the Familiarization Phase, pilots heard each of the 16 warning messages for their group at least 3, but not more than 30, times to give them an opportunity to understand and become familiar with as many messages as possible. The eight two-word and eight sentence-length messages were all presented on each of three familiarization runs without competing weather broadcast. Within a familiarization run, each message was repeated after 950 msec intervals until the pilot keyed his microphone and read back the message. After 10 repetitions of any 1 message, its presentation was automatically terminated under the assumption that the pilot was not going to understand it on that run. Following verbal message read-back, pilots wrote down whatever they had understood of each message on a prepared answer sheet. They were also asked to estimate the amount of time that they had spent listening to that particular message. The response time, which is the interval between the beginning of the first message repetition and the time when the pilot keyed his microphone, was recorded and used to calculate the number of message repeats and the ratio of estimated to actual time. Pilots were given no feedback on message comprehension or time estimation accuracy. This procedure generated the following data: a) response latency, b) number of message repeats, c) verbal time estimate, and d) pilots' written versions of the warning.

Recognition phase— During the Recognition Phase, pilots were asked to produce eight 10-sec intervals during which no other activity occurred (baseline) and then sixteen 10-sec intervals during which one of the eight two-word or eight sentence-length messages that had been presented during the Familiarization Phase was repeated one time only with a background of continuous weather broadcast. It was assumed that the majority of the messages had been comprehended during the Familiarization Phase and that for these familiar messages, the Recognition Phase of the experiment involved only simple recognition.

A pre-recorded message cued pilots when to begin each 10-sec production. By pressing one button, pilots initiated their production and the onset of the background weather broadcast. Approximately 2 sec later, 1 of the 16 warning messages was presented imbedded in the weather broadcast. Following warning-message presentation, which lasted no longer than 2.20 sec (table 1), the weather broadcast continued until pilots pressed a second button to terminate their 10-sec production. Then, they were asked to key their microphone and as quickly as possible, read back and then write down the warning message. This procedure generated the following data: a) length of time productions, b) pilots' written versions of the warning, and c) the length of time pilots waited after the end of their production to begin message readback (response latency).

RESULTS

Familiarization Phase

Overall performance change across runs—Intelligibility, as measured by the percentage of words correctly identified, increased significantly ($F = 5.12$; $df = 2,20$; $p < .025$) across the three familiarization runs (fig. 1(a)). The total time spent listening to each message (fig. 1(b)) decreased significantly ($F = 15.50$; $df = 2,20$; $p < .001$) across runs. The number of times that messages were repeated (fig. 1(d)) also decreased significantly ($F = 12.81$; $df = 2,20$; $p < .001$) across runs. The ratio of overall verbal estimate length to actual time spent listening to messages (fig. 1(c)) increased nonsignificantly ($F = 2.77$; $df = 2,20$; $p > .10$) across runs.

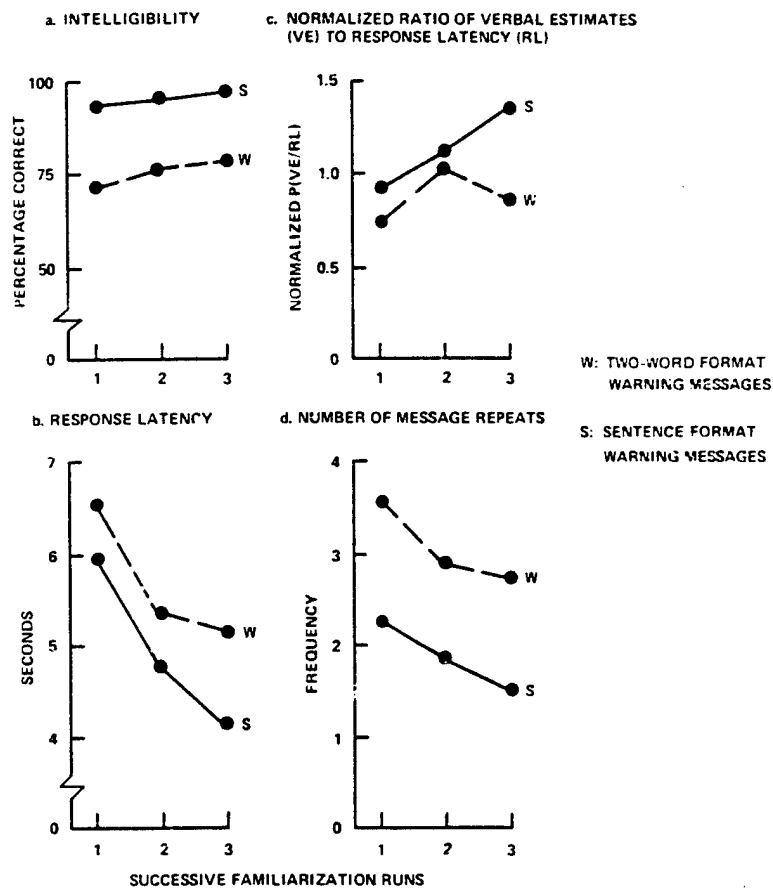


Figure 1.—Performance during the familiarization runs

Two-word versus sentence-length format—Warning messages presented in a two-word format were significantly ($F = 13.39$; $df = 1, 10$; $p < .005$) less intelligible than those presented in a sentence format for all runs (fig. 1(a)). Note that there were more words to be remembered in the sentence-length messages than the two-word messages. Even so, the sentence-length messages were more intelligible as a further consequence of the added linguistic redundancy.

Pilots allowed the two-word messages to repeat (fig. 1(d)) significantly ($F = 15.32$; $df = 1, 10$; $p < .005$) more often (at least one more time, on the average) than sentence-length messages. The time spent listening to two-word messages (fig. 1(b)) was also consistently longer than that spent listening to sentence-length messages even though each sentence-length message lasted longer than its corresponding two-word message. This difference was not statistically significant ($F = 2.06$; $df = 1, 10$; $p > .10$) because of a Pilot Group X Message Format interaction ($F = 7.09$; $df = 1, 10$; $p < .025$).

Verbal estimates of the length of time spent listening to two-word warnings were consistently shorter than the amount of time that had actually passed, whereas verbal estimates of intervals filled with messages presented in a sentence format were close to, or slightly longer than, elapsed clock time (fig. 1(c)). This difference in verbally estimated duration as a function of message format was statistically significant ($F = 25.79$; $df = 1, 10$; $p < .001$).

Influence of key word intelligibility—The 16 pairs of key words were rank ordered on the basis of intelligibility scores across familiarization runs, and then divided into quartiles that represented increasing intelligibility. As can be seen in figure 2(a), there was a large difference in intelligibility scores for two-word messages. Less intelligible two-word messages were repeated more often (fig. 2(b)) and were associated with the shortest verbal estimates (fig. 2(c)). These same key words, when presented in a sentence, were consistently more intelligible (fig. 2(a)). The facilitation provided by a sentence format was particularly great for the messages that were the least intelligible when presented in a two-word format. Sentences were repeated less often (fig. 2(b)) and sentences were associated with verbal estimates that most closely approximated the amount of time that had actually elapsed, for all quartiles of key word intelligibility (fig. 2(c)). The differences between two-word and sentence formats for all three measures (intelligibility, number of repetitions, and verbal estimate duration) were statistically significant ($p < .002$) in each case, as measured by Wilcoxon's signed ranks test for differences between 16 pairs of means.

Influence of message duration—Intelligibility was not correlated with simple message length. For example, the two-word message "DME Malfunctioning" was the longest two-word warning (1.5 sec), yet it was the least intelligible message when presented in a two-word format. The message "Your DME is Malfunctioning" was, by contrast, one of the most intelligible messages by virtue of the additional words "your" and "is" in the sentence context.

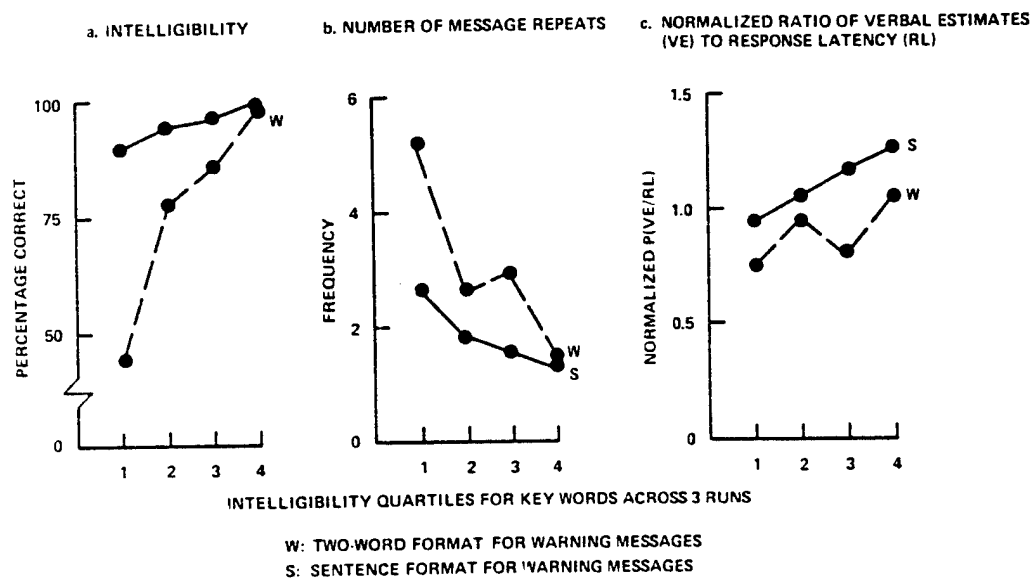


Figure 2.— Comparison of two-word versus sentence format for intelligibility quartiles (based on key word intelligibility for 3 familiarization runs)

Recognition Phase

During the one recognition run, the intelligibility scores were higher than they were during the three familiarization runs, even though each of the 16 messages was presented only once and with competing weather broadcast. Two-word message intelligibility increased from 76 percent to 82 percent and sentence intelligibility increased from 95 percent to 96 percent between the last familiarization run and the one recognition run.

Two-word versus sentence-length format— As before, sentences were more intelligible than two word messages (fig. 3(a)). Pilots also began sentence readback more quickly than two-word message readback (fig. 3(b)). Overall, produced durations lasted 12.55 sec, and were nearly identical for intervals filled with two-word and sentence-length messages (fig. 3(c)). This represents an increase of 2.5 sec in production length over the length of baseline productions made with no warning message presentation.

Influence of message familiarity— Several of the messages had not yet been completely understood at the end of the Familiarization Phase so 26 percent of the messages were still unfamiliar at the beginning of the Recognition Phase. This phenomenon occurred more often for two-word than sentence-length messages (fig. 4(a)). For messages that were still unfamiliar, response latencies were 1.0 sec slower (fig. 4(b)) and produced durations were 2.0 sec longer (fig. 4(c)) than were response latencies and produced durations for messages that had already been understood and were simply recognized. The length of produced durations that had been filled with unfamiliar two-word messages was 1.5 sec longer than the length of productions filled with

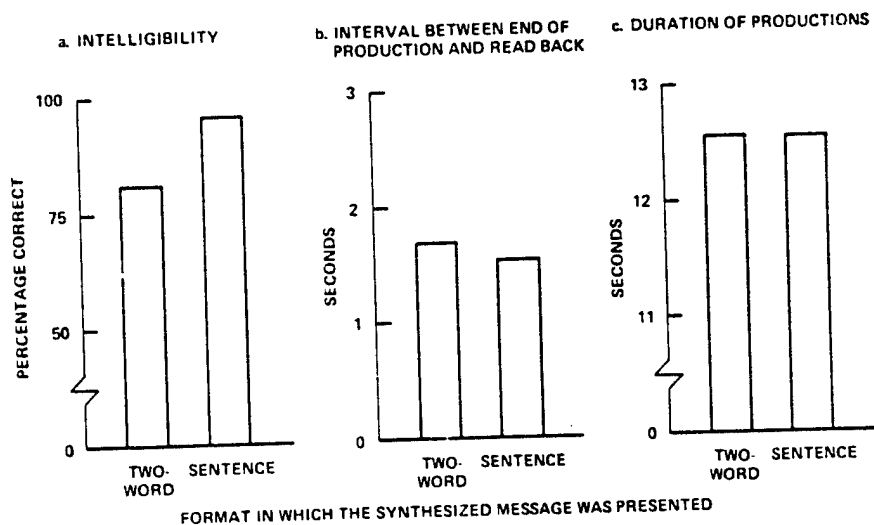


Figure 3.— Performance during the recognition phase

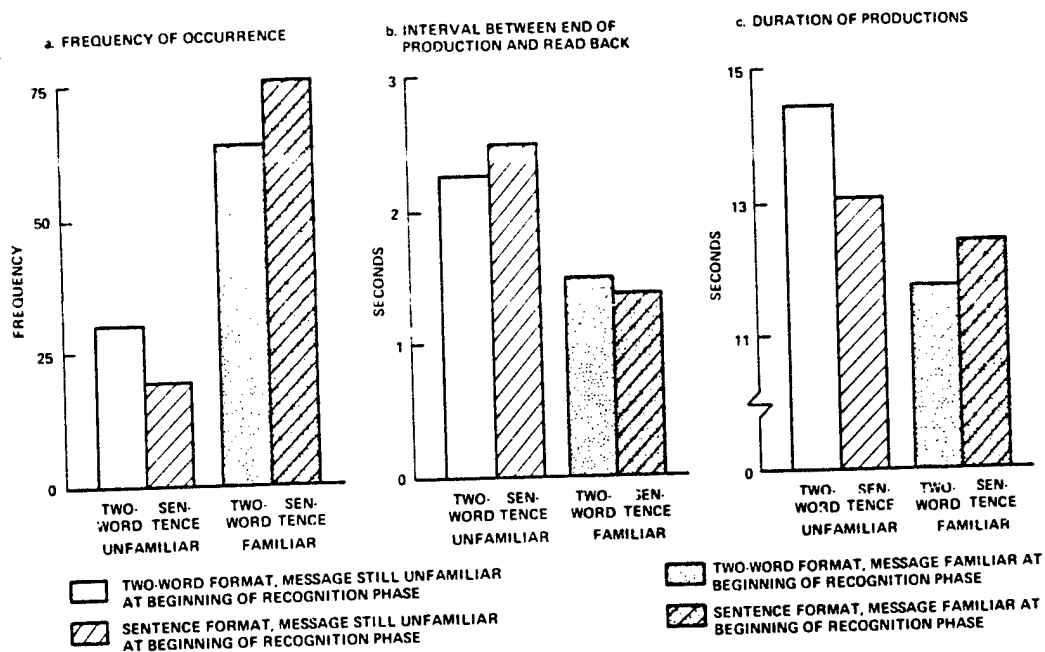


Figure 4.— Performance during the recognition phase as a function of warning message familiarity

unfamiliar sentence-length messages. For familiar messages, however, produced durations filled with two-word messages were slightly shorter (0.5 sec) than productions filled with sentences.

Influence of Number of Syllables

In order to compare the results of the present study to results by Simpson (ref. 3), the influence that the number of syllables had on intelligibility was also determined for both the Familiarization and Recognition Phases. While polysyllabic words were not necessarily longer in duration than monosyllabic words, the former were more intelligible, again suggesting that linguistic redundancy, rather than simple duration, was the most important factor in intelligibility. These results confirm the results of Simpson's research (ref. 3). During familiarization runs, both monosyllabic and

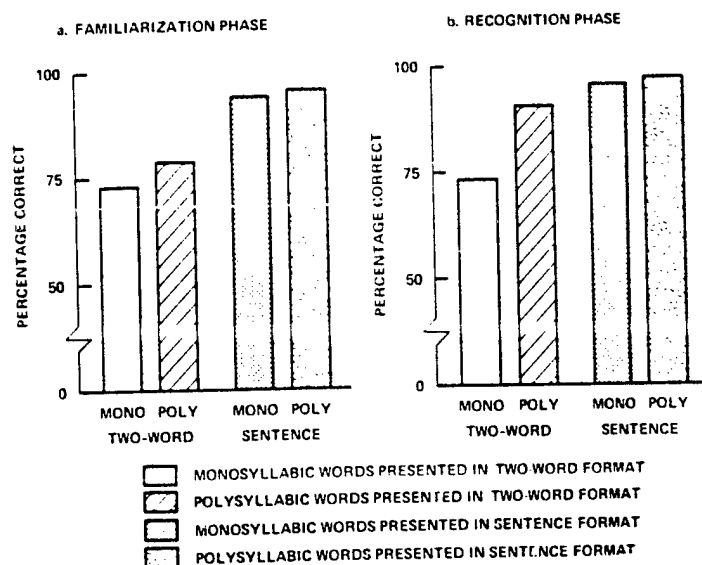


Figure 5.— The influence of monosyllabic and polysyllabic words on intelligibility as a function of message format

polysyllabic two-word warnings were less intelligible than the same monosyllabic and polysyllabic messages presented in sentences (fig. 5(a)). During recognition, with competing weather broadcast, polysyllabic words increased two-word message intelligibility to 91 percent. However, there was no corresponding gain in intelligibility for monosyllabic words (fig. 5(b)).

DISCUSSION

The results of the current research confirm and expand the research reported by Simpson (ref. 3) with respect to the facilitating influence of linguistic redundancy on initial comprehension and subsequent recognition. Warning messages presented in a sentence format were consistently more intelligible, required fewer repetitions and less listening time than two-word messages during familiarization, and were more easily recognized and read back more quickly during recognition.

Variation in the length of verbal time estimates and productions was used to assess the attention required to process messages differing in linguistic redundancy. It has been suggested that active verbal time estimation requires attention (ref. 1). If attention is diverted from timekeeping by concurrent activity, clock time continues whereas subjective timekeeping does not. This disparity results in an underestimation of elapsed time so that verbal estimates of the duration of an interval are too short. As expected, verbal estimates of the duration of intervals filled with less redundant, less intelligible, two-word messages were consistently shorter and less accurate than those of intervals filled with sentence-length messages during familiarization. These very short verbal estimates indicate that comprehending two-word messages may require more attention than comprehending similar messages presented in sentences.

It was further suggested (refs. 1, 2) that active time production also requires attention. If concurrent activity momentarily diverts attention from timekeeping, the amount of time that has passed may be underestimated causing subjects to wait too long to terminate their productions. This results in produced durations that are too long. In the present study, presenting weather broadcast and warning messages while the time production was in progress was associated with an increase of 2.5 sec in the average length of productions as compared to baseline. Produced durations filled with two-word messages were generally equivalent in length to those filled with sentences. This occurred even though sentence-length messages had the potential for greater distraction from timekeeping, simply because they lasted longer. If the foregoing hypothesis is correct, it appears that the linguistic redundancy provided by the sentence format not only facilitated understanding but did not require additional attention.

It is also interesting to note that produced durations filled with messages that had not yet been learned at the beginning of the recognition phase were 2.0 sec longer than those filled with previously learned messages that required only simple recognition. We infer that this length difference is a consequence of the additional attention required to comprehend a message that had not yet been learned as compared to the relatively simple task of recognizing a previously learned message. This difference in produced durations was particularly great for two-word messages which appear to be less intelligible and so require more attention when unfamiliar than when familiar.

Sentence-length warnings appear to require less attention to comprehend initially than two-word messages. The former were associated with productions that were 1.5 sec shorter than those for intervals filled with two-word messages. However, sentences that had been previously learned had longer production times than did two-word messages that were familiar. Once a familiar message had been recognized, pilots had to hold it in memory during the remainder of the production. Distraction from time estimation most likely varied as a function of the number of items held momentarily in memory. If pilots attended to the passage of time only at the end of each mental repetition of the recognized message, the sentences, being longer units, resulted in less frequent attention to time, and thereby longer productions, than did the two-word warnings.

CONCLUDING REMARKS

From these, as well as previously reported results, it appears that verbal and produced time estimates may provide a useful index of the attention demands of concurrent activity. Factors associated with the intelligibility of synthesized warning messages, such as redundancy and familiarity, presumably influenced the attention required to process them. The length of verbal estimates and produced durations provided an indirect measure of the amount of attention required by messages differing in intelligibility.

The results of this research, and of that reported by Simpson (ref. 3), have demonstrated the facilitating influence of linguistic redundancy, such as that provided by a sentence format, on cockpit-warning comprehension and recognition. Even though sentences last longer and contain more words to be remembered than do two-word messages, the former require fewer repetitions, less time, and less attention to comprehend initially and are read back more quickly and accurately once they are familiar.

REFERENCES

1. Hart, S. G.: Time Estimation as a Secondary Task to Measure Workload. Proceedings of the 11th Annual Conference on Manual Control. NASA TM X-62,464, Ames Research Center, Moffett Field, California, 1975.
2. Hart, S. G., and McPherson, D.: Airline Pilot Time Estimation During Concurrent Activity Including Simulated Flight. Paper presented at the 47th Annual Meeting of the Aerospace Medical Association, Bal Harbour, Florida, May, 1976.
3. Simpson, C. A.: Effects of Linguistic Redundancy on Pilot's Comprehension of Synthesized Speech. Paper presented at the 12th Annual Conference on Manual Control, Champaign, Illinois, May, 1976.
4. Miller, G. A.; Heise, G. A.; and Lichten, W.: The Intelligibility of Speech as a Function of the Context of the Test Materials. *Journal of Experimental Psychology*, vol. 41, 1951, pp. 329-335.

TACTUAL COMMANDS FOR PILOT FLARE TRAINING*

Richard D. Gilson and Ronald W. Ventola

Departments of Aviation, Psychology
and Electrical Engineering

The Ohio State University
Columbus, Ohio

INTRODUCTION

The problem of aircraft control during approach and landing operations deserves strong attention, as evidenced by aircraft accident statistics. Approximately one-half of all aircraft accidents occur during such operations despite their contributing only a brief portion of the total flight time. (1)

A primary cause is the heavy demand placed on a pilot--especially the division of visual attention required to control the flight path and airspeed. During the approach to landing, information pertaining to the flight path is primarily obtained from visual cues outside the cockpit, while pitch information is obtained via a panel-mounted airspeed display.

In addition, just prior to touchdown, the pilot's visual attention is progressively drawn towards runway-specific cues allowing little, if any use of instrument panel information.

It is hypothesized then, that the use of a non-visual display would allow:

- (a) An alleviation of this unwanted division of visual attention, and
- (b) the opportunity of presenting accurate pitch-command information during the roundout (or flare) just prior to touchdown.

* The efforts reported here were sponsored by the National Aeronautics and Space Administration (NASA), Contract No. NAS 2-8954.

this should be especially beneficial to a novice pilot - who lacks the experience to use relevant visual, inertial, and aural cues. This should result in the following benefits:

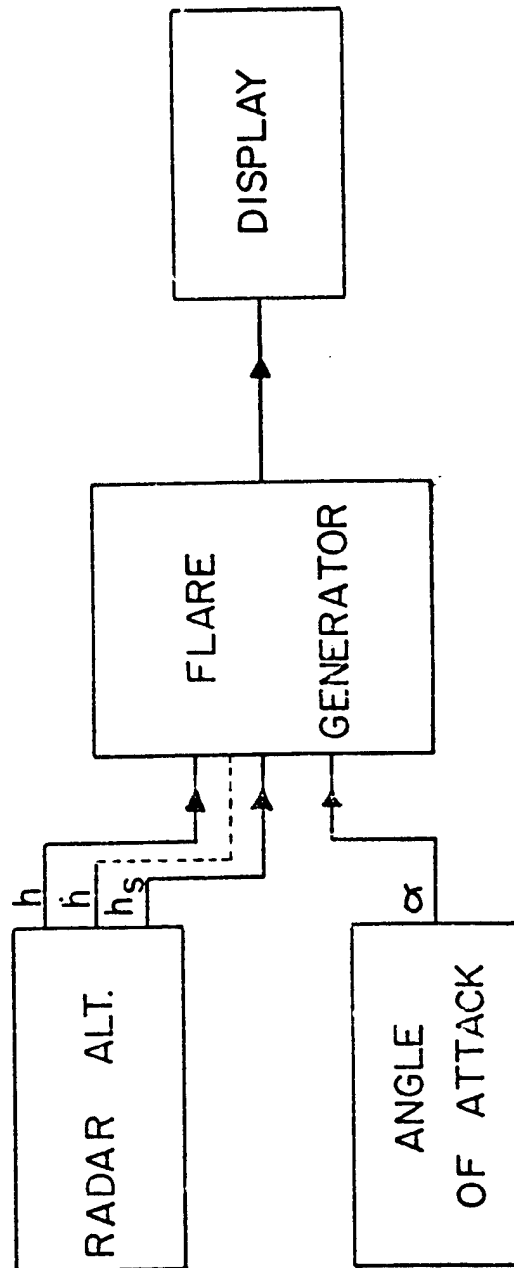
- (1) His task should be simplified
- (2) His performance should be improved and
- (3) His learning of the appropriate visual cues from outside the cockpit should be enhanced.

Previous work by Fenton, Gilson and others (2, 3, 4 & 5) has shown an effective and practically implemented non-visual display for automobile, aircraft, and helicopter controls. Based on their findings, the use of a kinesthetic-tactual display appears to be a natural way of overcoming the unwanted division of visual attention during flight, while also avoiding any further division of auditory attention.

In the present study, information pertaining to the aerodynamic state of an aircraft was continuously presented via a kinesthetic-tactual display during takeoffs and landings. Figure 1 depicts the desired control information during an approach and landing. The tactual display is programmed to present critical-pitch directives from the beginning of takeoff roll through the approach and flare-to-landing. The reference input is the desired angle-of-attack (AOA) which is, of course, related to the desired approach airspeed. Whenever the aircraft is higher than 50 feet above the ground the desired angle-of-attack is held constant. Below 50 feet the desired AOA is an increasing inverse function of height, so that the aircraft is near stall at liftoff and at touchdown.

Figure 2 shows the control loop. The feedback signal is the measured AOA. The display input is the difference between the measured and reference (or desired) AOA--thus, a compensatory tactual "pitch director."

The tactual display was built into the head of a conventional aircraft control yoke. The display itself, photographed in Figure 3 consisted of a moveable section of the control grip. This is shown as protruding from the forward part of the grip and recessed into the aft part. This protrusion corresponds to an unwanted increase in angle-of-attack, and a pilot responds by moving the yoke forward so as to decrease this angle and return the display to its neutral or flush position. Next, in Figure 4 is a view of the display protruding "backward" which requires an aft corrective motion of the control yoke. In essence, the pilot



h = Absolute Altitude
 \dot{h} = Vertical Velocity
 h_s = Flare Altitude set
 α = Angle of Attack

Figure 1. Desired Control Information

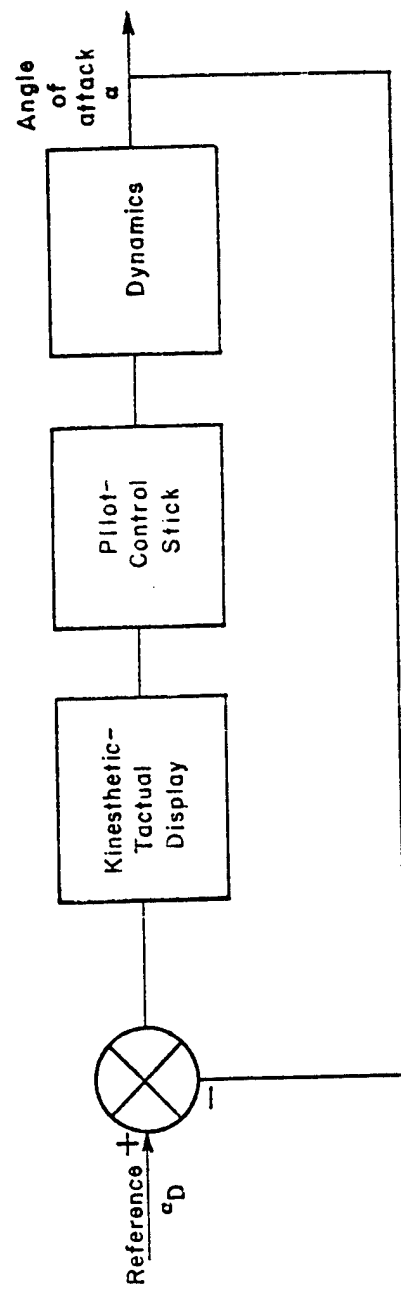


Fig. 2 Control loop for angle-of-attack



Figure 3 Kinesthetic-tactual display
protruding forward



Figure 4 Kinesthetic-tactual display
protruding rearward

ORIGINAL PAGE IS
OF POOR QUALITY

follows the display commands to reduce errors to zero.

PROCEDURE

The experimental inflight study considered novice pilot performance while flying approaches to landing. Initially, each of 12 novice pilots received flight instruction with a standard, 3-hour primary training syllabus designed to teach them fundamentals of aircraft control (but without allowing practice of landings.) At the completion of this Preliminary Training, the subjects were evaluated on a standard series of test maneuvers designed to equally match subjects, by performance, into two groups.

The first experimental phase then commenced, wherein the subjects flew four one-hour test periods of six takeoffs and landings. One group, designated the airspeed group, received only the conventional visual display of airspeed. The other, the tactual group, received only the tactual displayed information.

The second experimental phase then followed, consisting of two one-hour test periods, wherein the display conditions were reversed: that is, the airspeed trained group now flew with the tactual display, while the tactual trained group received visual airspeed information only. In all the experimental phases, performance was analyzed during the final approach and landing from approximately $1\frac{1}{2}$ miles out to touchdown. Performance measures included: the number of instructor pilot (a) verbal assists (b) assisted landings and (c) actual takeovers during landing (where safety of flight was involved). Touchdown performance measures consisted of: (a) lateral deviations from runway centerline, (b) longitudinal touchdown position measured from the fixed distance marker and (c) sink rate or "g" loading at touchdown. Other performance measures recorded but not as yet analyzed include, angle of attack, flight controls positions, runway lineup, and glide path errors during the final approach.

RESULTS AND DISCUSSION

Preliminary data are shown in Figures 5 and 6. Figure 5 depicts the total number of unassisted landings performed by

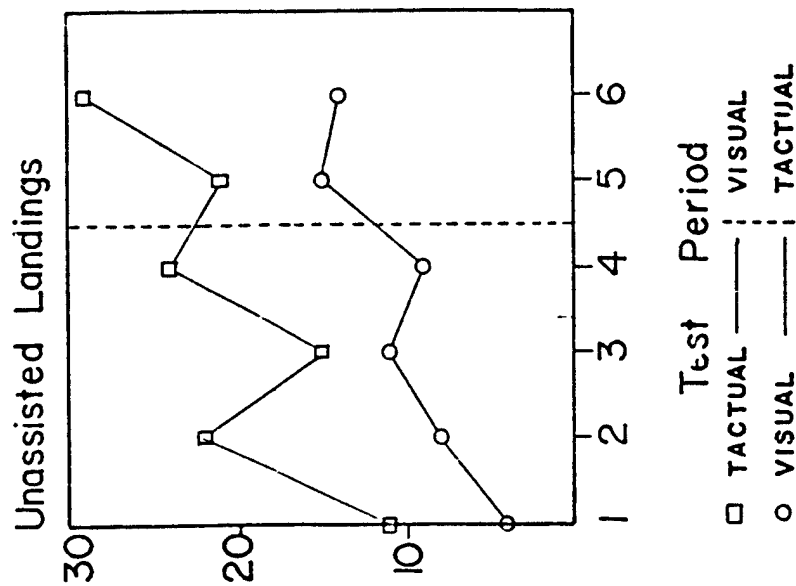


Figure 5.
Total number of
Unassisted Landings
36 possible/group/period

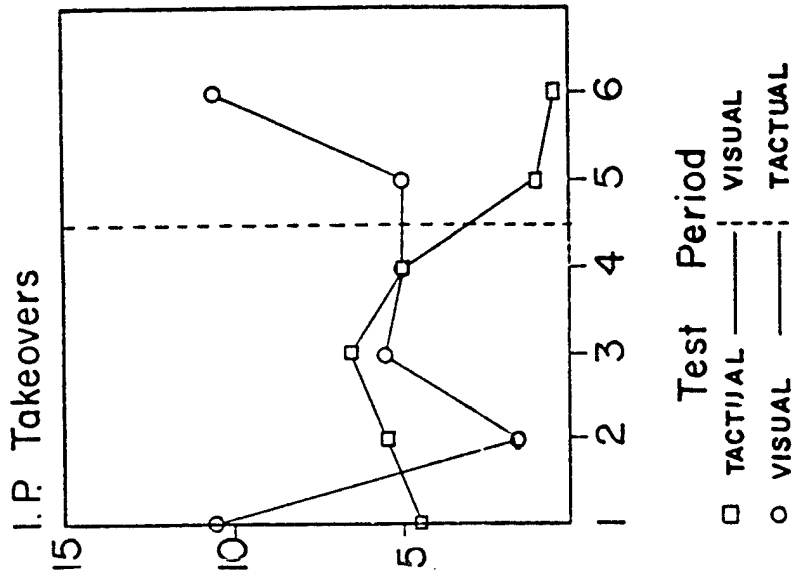


Figure 6.
Total number of
Instructor-pilot takeovers
36 possible/group/period

subject groups per test period. Unassisted landings were those expected not to exceed three "g's" at touchdown. With 6 landings per hour and 6 subjects per group, a total of 36 landings per test period were possible. The group initially trained with the tactual display significantly ($p < .01$) out-performed the airspeed group not only in the first four periods but subsequently when they were transferred to fly with the visual airspeed indicator only.

This suggests that the group initially trained with the tactual display was facilitated in their flares-to-landing by following the display (50% unassisted landings), and that they had learned enough to subsequently perform quite well (almost 70% unassisted landings) with the conventional airspeed display only. The latter rejected an initial hypothesis that a "crutch-like" dependency on the display might develop when this group was transferred to the visual-only condition. On the other hand, the group first trained with the airspeed display performed like typical flight students with approximately 22% unassisted landings during the first four periods. What was remarkable, however, was that the airspeed group did not significantly improve when they were transferred to the tactual display-- there were only 40% unassisted landings in the last test periods (5 & 6) compared to 50% unassisted landings in the initial training periods for the tactual group.

The reason for this lack in improvement is suggested in Figure 6--summarizing the number of instructor takeovers. An analysis of variance for the takeover results shows a significant interaction effect of group by test period ($p < .05$).

It is apparent that when the airspeed group was transferred to use the tactual display (shown by test periods 5 & 6) there was a dramatic increase in the number of takeovers. These data indicate that some apparent conflict occurred when the airspeed group was transferred to the use of the tactual display. This conflict may occur for the airspeed group because their initial strategy, in utilizing the available visual cues, developed largely by trial and error. Such strategies most likely differed from the approach commanded by the tactual display. Conversely, the group initially trained with the tactual display was forced to use and observe the desired flare-to-landing strategy. Thus, when this tactual group was transferred to use the visual airspeed display, they relied on what they had learned before rather than trial and error.

Among the other performance measures, no significant differences were found in either the number of verbal assists given by the instructor-pilot, or in the vertical velocity or "g" loading, just at touchdown for unassisted landings. This suggests that no apparent biases were introduced by the instructor-pilot.

CONCLUSIONS

There are several directions of future research that are envisioned at this time.

(1) For pilot training purposes, tactual commands may benefit a student's understanding of the interactions between aircraft attitude, speed and power by means of a tight control loop and the display's natural stimulus-response compatibility.

(2) For more basic studies, cognitive, perceptual, and motor learning may be investigated by employing appropriate tactual feedback to differentially augment visual or tactual display-control relationships. Such studies are currently being initiated in a flight simulator with a computer graphics display of the landing environment.

(3) For practical aircraft control, kinesthetic-tactual displays incorporated within the appropriate control levers could allow for zero-zero landings. For example, during an instrument approach, a power command could be located in the throttle lever and a lineup command could be incorporated in a multiaxis control yoke. Indeed, the tactual pitch commands currently allow an ILS approach to touchdown if the cross-pointer indicator is utilized.

In summary, the exploration of the efficacy of tactual displays in the aircraft environment has only begun and many, perhaps significant, improvements may be made in both the safety and precise control of aircraft.

REFERENCES

- (1) National Transportation Safety Board, Annual Review of U.S. General Aviation Accidents Occuring in Calender Year 1969.
- (2) Rule, R.G., and Fenton, R.E. On the Effects of State Information on Driver-Vehicle Performance in Car Following, IEEE Transactions SMC, No. 5, November 1972, pp.630-637.
- (3) Gilson, R.D. and Fenton, R.F. Kinesthetic-tactual information presentations -- Inflight studies. IEEE Transactions SMC, No. 6, November 1974, pp. 531-535.
- (4) Gilson, R.D. and Fenton, R.F. Development of a Concept for a Stall-Deterrent Device for Small Airplanes, Federal Aviation Administration, Report No. FAA RD-75-53 June 1975 NTIS Springfield,VA
- (5) Dunn, R.S. and Gilson, R.D. A Simulation Study of Helicopter Pilot Workload Reduction Using a Tactile Display. (See These Proceedings).

TWO-DIMENSIONAL COMPENSATORY TRACKING WITH TACTILE DISPLAYS*

By H. P. Schmid
General Dynamics Pomona Division, Ph.D. Fellow
University of Southern California

G. A. Bekey
Department of Electrical Engineering
University of Southern California,
and Rancho Los Amigos Hospital

J. B. Reswick
Rehabilitation Engineering Center,
Rancho Los Amigos Hospital and
University of Southern California

SUMMARY

Compensatory tracking experiments in one and two axes using tactile displays are described. The electro-tactile inputs to the operator are located circumferentially about the waist region utilizing four electrodes. The error signal, applied to the operator through the electrodes, is an amplitude modulated 100 Hz pulse train with a fixed pulse width of 40 μ sec. The dynamic range of the display between sensation threshold and pain is one order of magnitude. The operator's output is isometric muscle force applied to a two degree of freedom joystick. Describing functions for both one and two-dimensional tracking with low frequency pseudo-random noise inputs are presented. The results indicate that tactile tracking with this particular display in one- and two-dimensions is indeed practical. Coherence analysis of the two-dimensional tracking task indicates that, for the low frequency forcing function utilized, no significant linear cross-coupling exists in the operator model. Therefore, a two-dimensional task can be considered to be two independent one-dimensional tasks. A comparison of the tracking ability of traumatic paraplegics with that of normal individuals allows the conclusion that there are no essential differences between these groups.

INTRODUCTION

Paraplegics with spinal cord lesions at or above L3 have a great deal of difficulty in ambulating. One of their major problems is the lack of motor control over knee and hip joints. Poliomyelitis paraplegics, whose condition is due to a virus which selectively attacks motor fibers, have their sensory ability remaining intact, which allows them to perceive the position and velocity of their paralyzed limbs. This perception of position and motion is called proprioception. In contrast, traumatic

* Supported in part by the Rehabilitation Services Administration, DHEW under grant number RD-23-P-55442.

paraplegics with a spinal cord lesion due to a traumatic injury usually have neither muscular control nor proprioception of the affected limbs. The poliomyelitis paraplegic with no hip control can functionally walk in long leg braces with forearm crutches, whereas the traumatic paraplegic chooses to utilize a wheelchair. The lack of proprioception in the legs and hips is the only difference between these individuals. Based on this fact, it has been hypothesized that the provision of synthetic sensory feedback of body posture information from the affected areas will help the traumatic paraplegic to maintain a better static vertical posture. The maintenance of static posture is a prerequisite for stable, possibly electrically stimulated, gait.

Out of the three commonly used channels of communication, which are visual, auditory and tactile, the latter was chosen to be the mode of information display for the sensory feedback. The visual channel is utilized extensively during walking, mostly for path choice in obstacle avoidance and for the admiration of scenery. Although the auditory channel is less utilized in a walking task, it can be most easily jammed or overloaded by inputs such as horns, sirens, etc. For both visual and auditory displays the problem of providing a cosmetically acceptable mechanization presents great difficulty. For the tactile channel, especially when the display is located in the waist region, the problems of interference and cosmetic acceptability are greatly diminished.

A number of tactile tracking studies have been reported; their approach and results are summarized in Reference 1. The displays for these studies were mainly vibratory. They were applied at diverse body locations such as on the thumbs, fingernails, hands, arms, forehead and chest. All studies concluded that tactile tracking is possible, but with generally poorer performance as compared with visual displays. More recently, Reference 2 reports experiments with tactile displays for aircraft control. In this study, electrotactile and vibratory displays were utilized. Better consistency and pilot acceptance were attained with the vibratory display. The displays were multi-electrode or vibrator type arranged in + or H fashion on the chest and abdomen of the pilot. The problems with all previous studies are that they do not provide an easy mode of presenting body posture information to the subject wearing the display. The models presented also were not analyzed for cross-coupling components and do not provide a model of the human tracking an input whose power spectrum resembles static posture sway. Humans standing erect control small oscillatory movements about the normal upright position; this deviation is called sway. It has been measured and its power spectral content has been reported in Reference 3.

The basic objectives of this study are (1) to show the feasibility of tracking a forcing function, having a power spectral density (PSD) similar to standing sway in humans, utilizing an electro tactile display of four electrodes located about the waist, (2) to model the human operator tracking this forcing function in one- and two-dimensional compensatory tasks, (3) to investigate the cross coupling terms in the human operator model in order to determine their relative importance and (4) to compare the tracking ability of normals to that of paraplegics.

EXPERIMENTAL DESIGN

The simple tactile compensatory tracking task considered in this experiment is shown in Figure 1. The subject in the loop has the task of minimizing the error displayed to him. A pseudo-random binary sequence generator, appropriately filtered, provides the forcing function c to be tracked. The subject's force response r , a function of his voluntary muscle force transformed by the isometric joystick manipulator, is subtracted from c to provide a measure of the tracking error e . The tactile display signal generator produces negative constant current pulses at a frequency of 100 Hz with a duration of 40 μ sec and an amplitude directly proportional to the error. The tactile display consists of an elastic belt, four (4) electrodes and the associated wiring shown in Figure 2. The placement of the electrodes about the waist region is shown in Figure 3. Details of the surface electrode can be seen in Figure 4. These electrodes are constructed from silver according to the findings in Reference 4 and are of the concentric type. The center circular region is the active area supplying the negative current. It is isolated from the rectangular ground plane by a small teflon spacer.

The dynamic range of the display is one order of magnitude between the threshold of sensation and the threshold of pain. This range is indicated in the strength duration curves presented in Figure 5. There the current stimulus amplitude for sensation and pain threshold is plotted for a number of pulse durations. The duration of 40 μ sec was chosen since it provides a reasonable dynamic range without requiring unduly large voltages. The electrodes are applied to the skin after a small amount of electrode gel is applied. The electrodes in this configuration present a dynamic impedance of approximately 2-4K ohms. The minimum level of stimulation throughout the experiments was adjusted to the threshold of sensation in order to eliminate the dead zone inherent in this display configuration.

The two-dimensional tracking system is shown in Figure 6, in block diagram form. The pseudo-random noise generator is a 17 stage digital shift-register with a modulo-2 adder in the feedback path. Design details may be found in Reference 5. An original sequence S_1 and a delayed version S_2 can be generated through utilization of additional modulo-2 adders. For this particular design, utilizing a clock rate of 30 Hz, the duration of the sequence is 72.8 minutes. The delayed sequence is shifted by 36.4 minutes providing two independent signals. Each of these sequences are filtered by identical filters labeled A/P (anterior/posterior) and Lat. (lateral) sway filter. The output of each summing junction, representing the tracking errors e , is amplified, separated into positive and negative components, synchronously sampled, converted to stimulus current and finally applied to the subject via the electrodes. The tactile display gain (TDG) was set at five for all experiments presented and is included in the models. This gain setting minimizes the tracking error; that is, lower and higher gains tend to increase the tracking error score as measured by the performance monitor (PM). The performance monitor calculates the integral of the absolute value of the tracking error, which is displayed on a meter for on-line performance evaluation. The isometric force response measured by

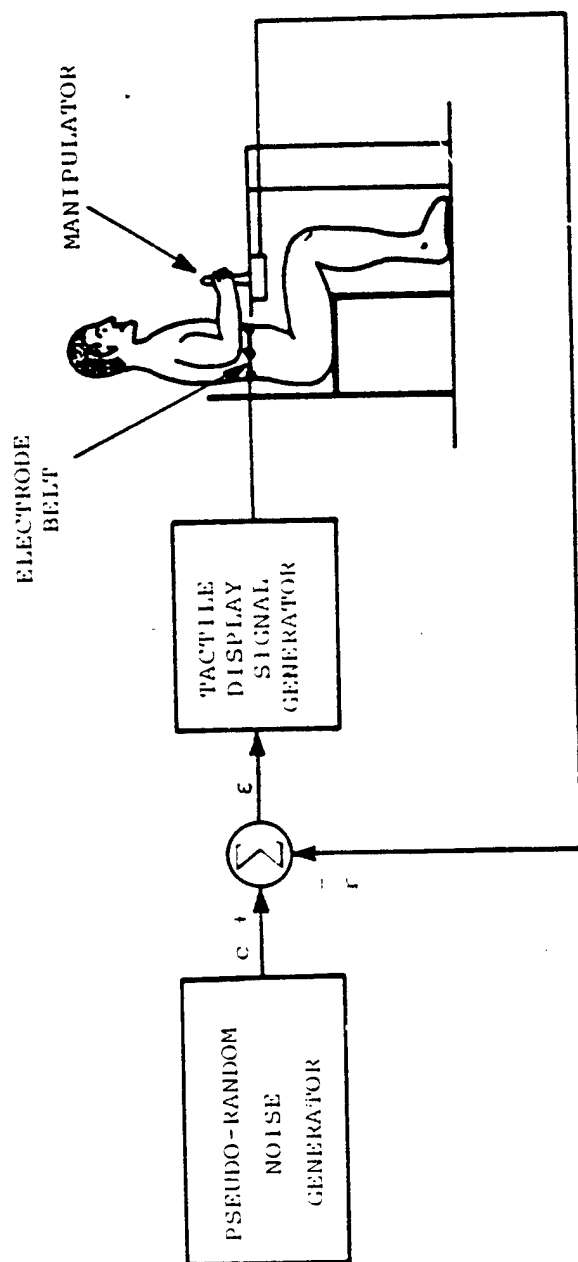


FIGURE 1. Tactile Compensatory Tracking Task

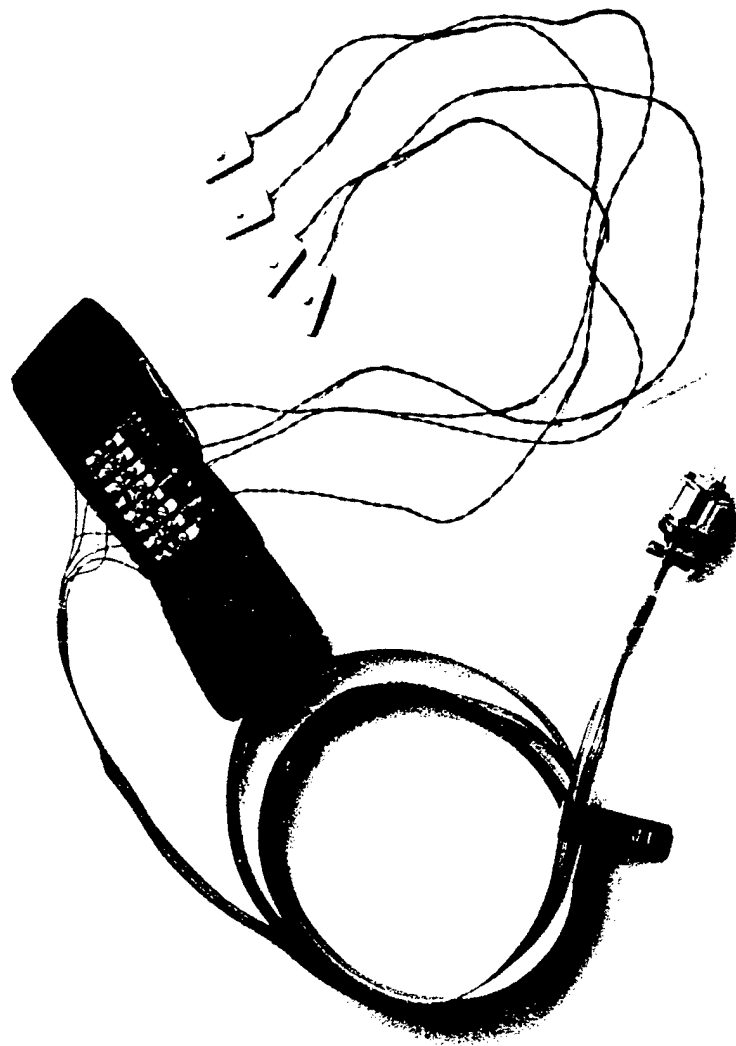


FIGURE 2. Tactile Display (Electrode Belt).

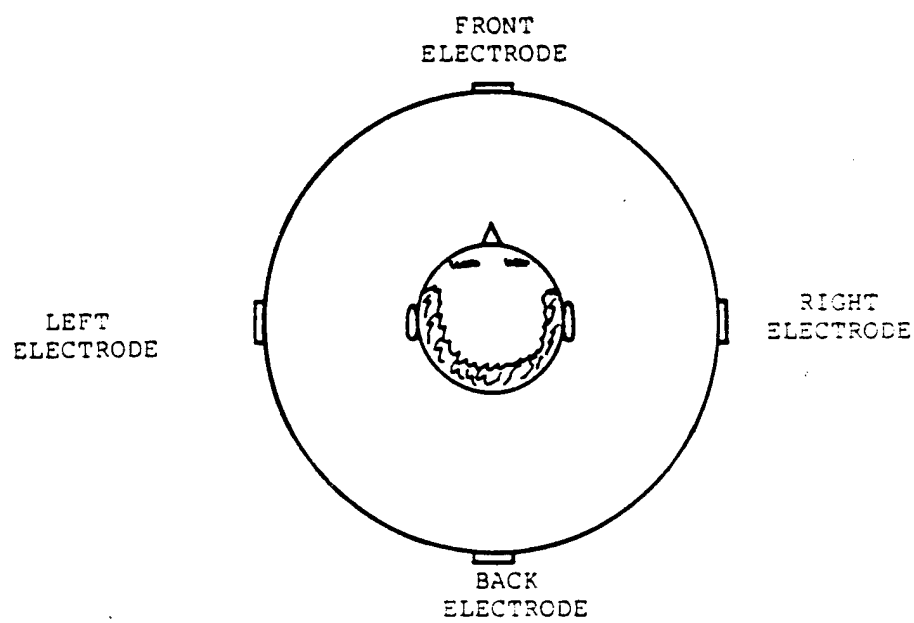


FIGURE 3. Electrode Placement

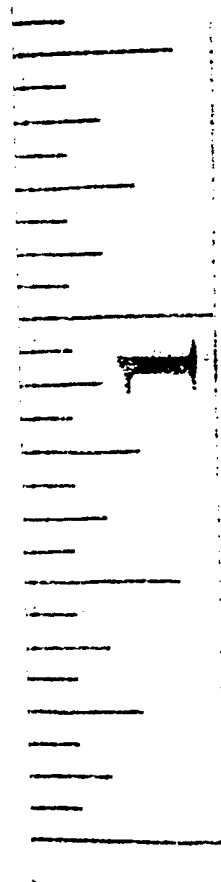
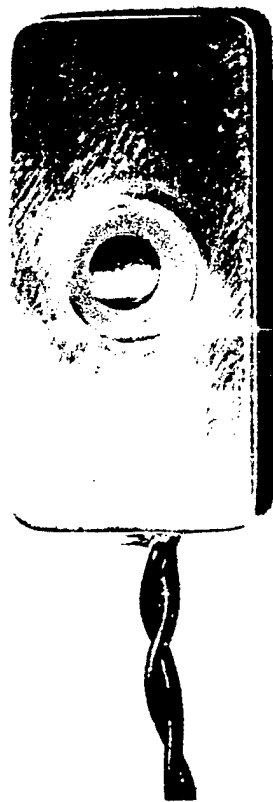


FIGURE 4. Surface Electrode

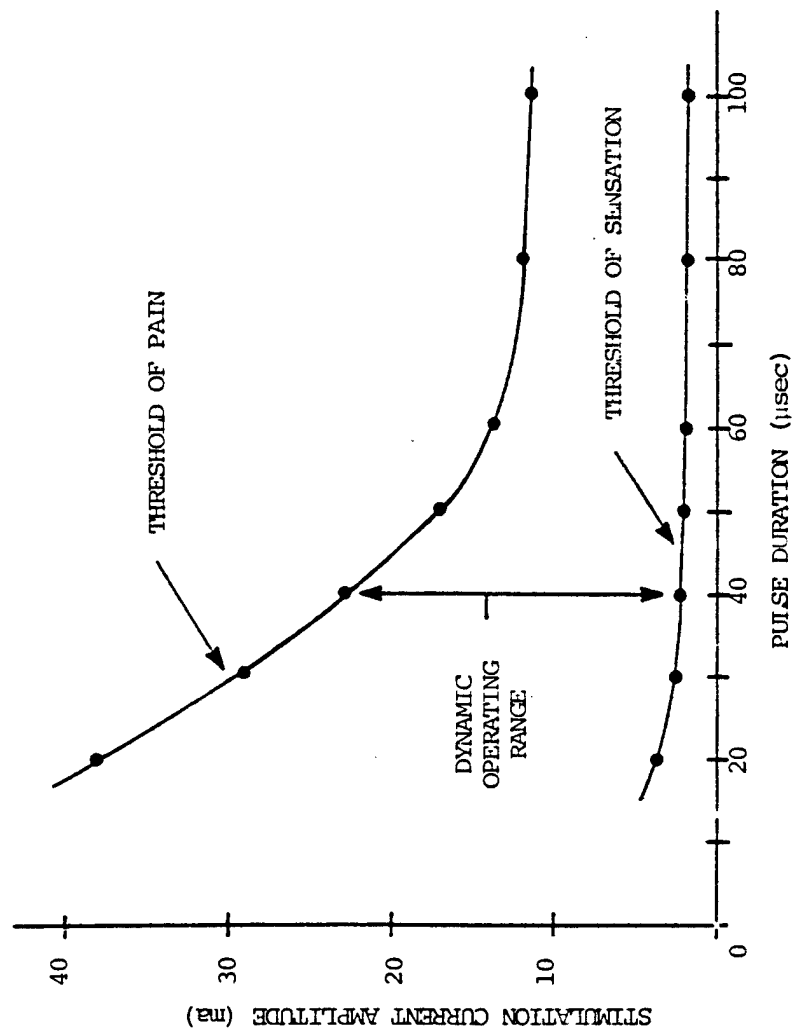


FIGURE 5. Strength Duration Curve Indicating Dynamic Range of Display

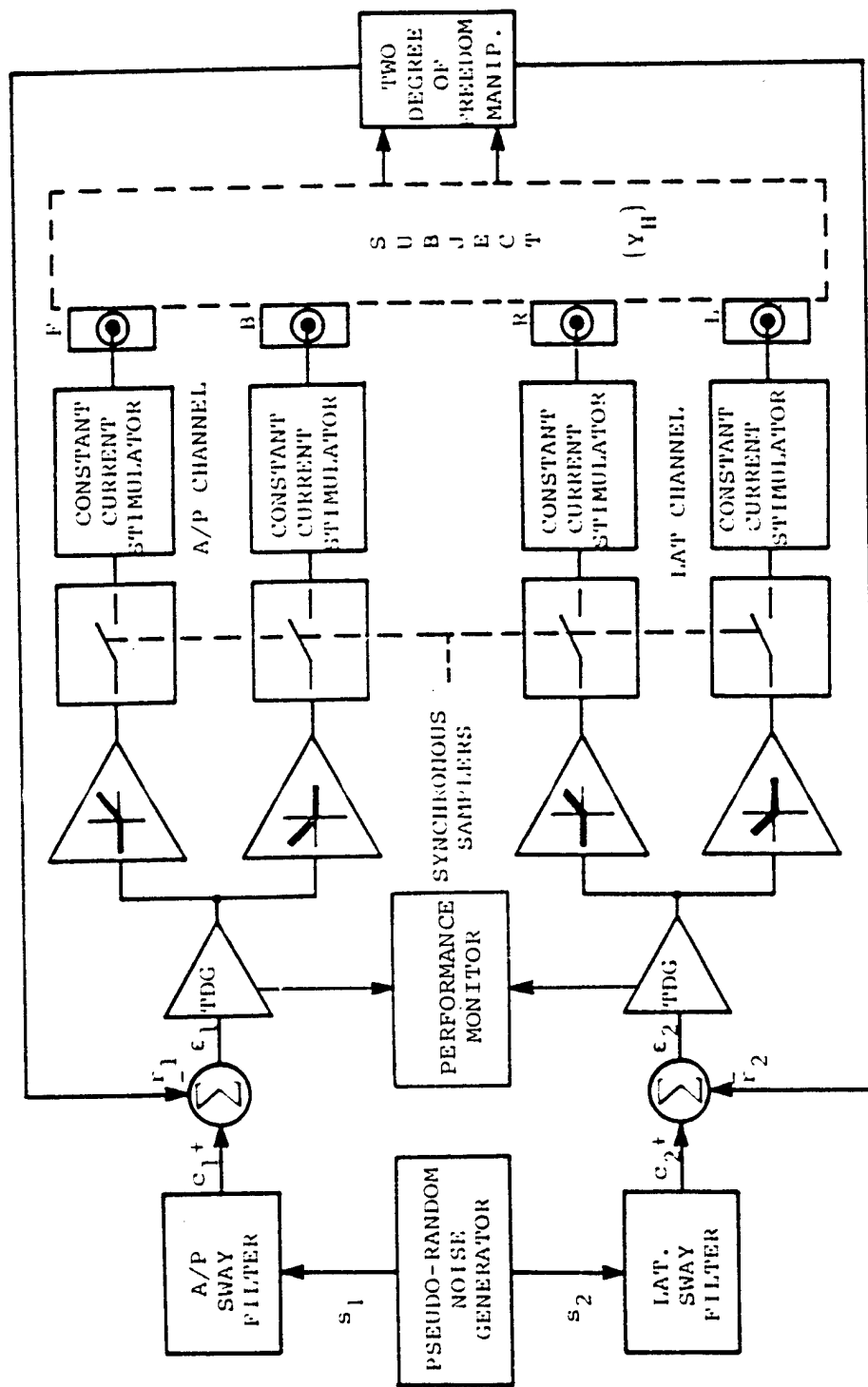


FIGURE 6. Tracking System Block Diagram

the two degree of freedom manipulator (joystick) is fed back to the summing junction.

The physical configuration of the tracking system hardware is shown in Figure 7 and contains all the elements described above.

One paraplegic and three normal subjects have been tested in two one-dimensional and one two-dimensional compensatory tracking task. The one-dimensional tasks are (1) tracking in the anterior/posterior plane and (2) tracking in the lateral plane. The two-dimensional task is the combination of these individual tracking tasks.

For each experiment the forcing functions (c_1 and c_2), the tracking error (e_1 and e_2), the force response (r_1 and r_2), and a 100 Hz trigger signal synchronous with the stimulation rate are recorded on analog tape. The analog data is later digitized at the 100 Hz rate and finally analyzed at a 10 Hz rate. Proper precautions are taken to reduce aliasing due to sampling by appropriately filtering the data at each step of the operation.

The time history of the recorded signals for a typical tracking run are shown in Figure 8. The similarity between the forcing function and the force response are obvious. The tracking error is plotted on an expanded scale.

ANALYSIS

Quasilinear describing functions, based on spectral analysis, are utilized to represent the human operator model in this tactile compensatory tracking task. The model for the human operator in a single axis task is shown in Figure 9. Models of this form have been extensively discussed in the literature on human operators, such as Reference 6. The open loop describing function for the human operator (Reference 6) is found to be

$$Y_H(j\omega) = \frac{S_{cr}(j\omega)}{S_{ce}(j\omega)}, \quad (1)$$

where $S_{cr}(j\omega)$ and $S_{ce}(j\omega)$ are the input-output and input-error cross-power spectral densities.

Estimates of the cross-power spectral densities are obtained by a method of averaging modified periodograms as described in Reference 7. In this procedure the time domain data is sectioned into M overlapping segments of length N . The overlap here has been chosen to be $N/2$ in order to minimize the variance of the spectral estimate for the total available data length. Each data section is tapered in some fashion; the Hamming Window was chosen to perform this function. The Fourier Transform, defined as

$$X_k = \sum_{n=0}^{N-1} x_n w_n e^{-j \frac{2\pi nk}{N}} \quad k=0, 1, \dots, N-1 \quad (2)$$

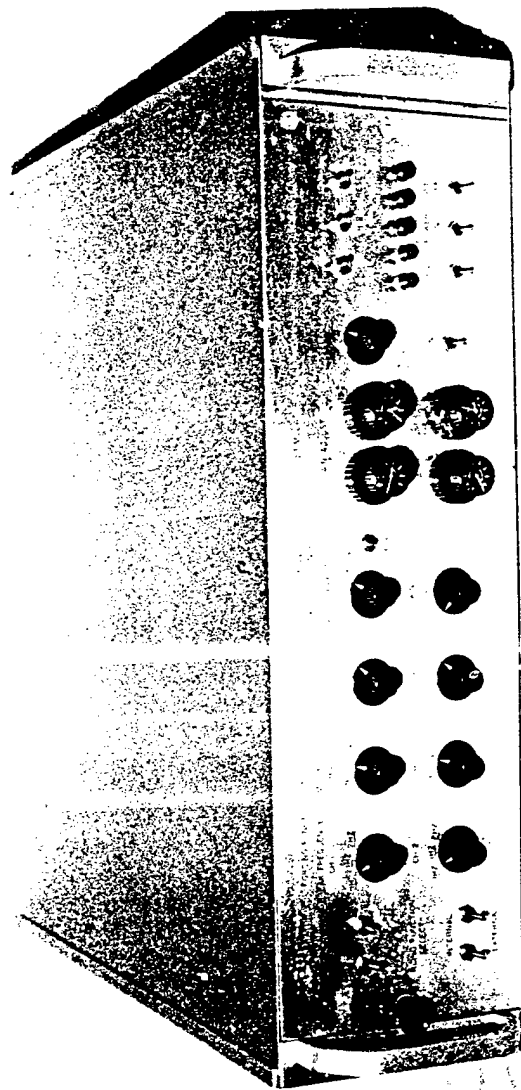


FIGURE 7. Tracking System Hardware

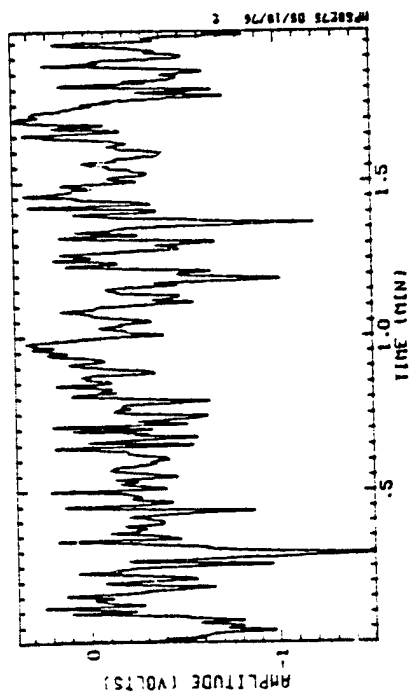
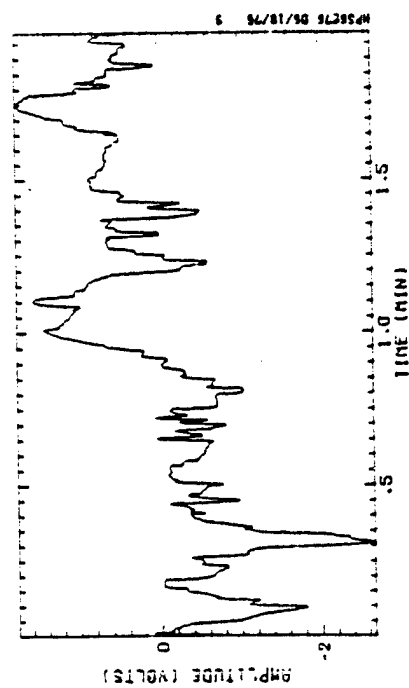
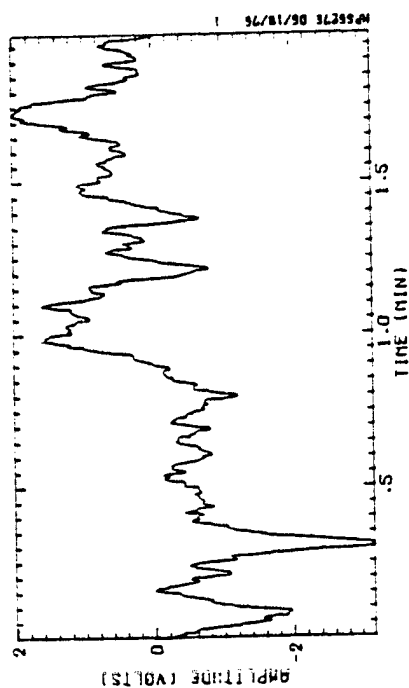


FIGURE 8. Typical Tracking Signal Time History

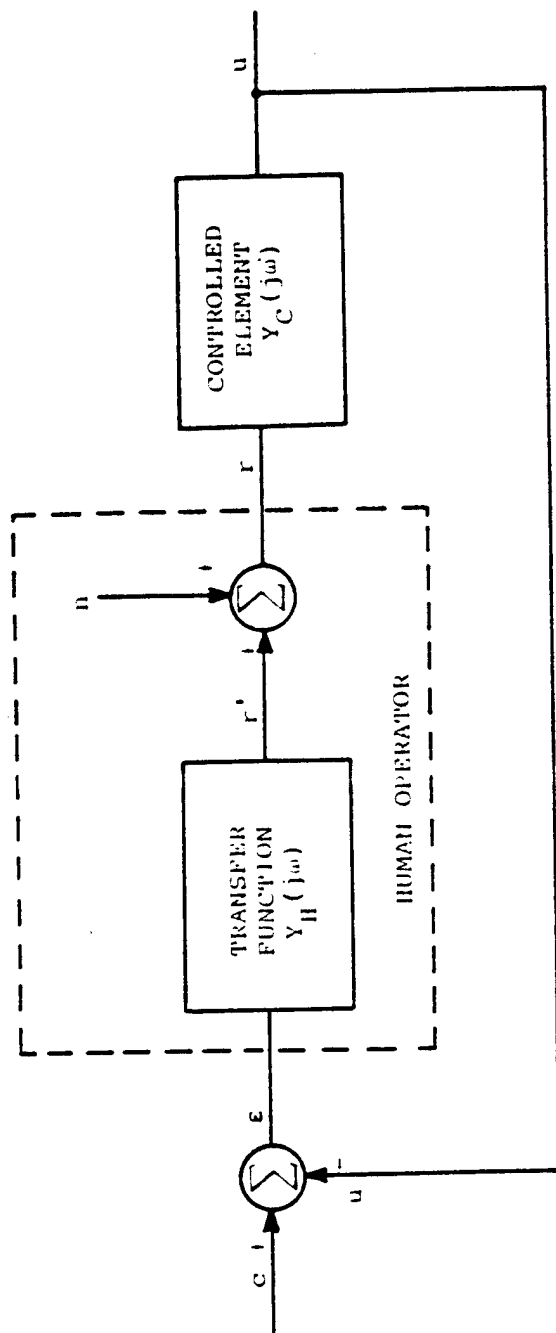


FIGURE 9. Single Axis Compensatory Tracking Task

is calculated for each windowed section utilizing a standard FFT algorithm. The modified periodogram

$$I_m(w_k) = \frac{\Delta t}{NU} |X_k|^2 \quad m=1, \dots, M \quad (3)$$

is formed, where U is a function of the total energy in the window and Δt is the time between samples. The estimate of the Spectral Density is then obtained by averaging the modified periodograms according to the equation

$$S_{xx}(w) = \frac{1}{M} \sum_{m=1}^M I_m(w_k) \quad (4)$$

Cross-power spectral density estimates are obtained by following the same procedure, except in this case the modified periodogram is defined as

$$I_m(w_k) = \frac{\Delta t}{NU} |X_k^* Y_k| \quad (5)$$

where X and Y are the Fourier transforms of the two signals under consideration and the asterisk indicates the complex conjugate.

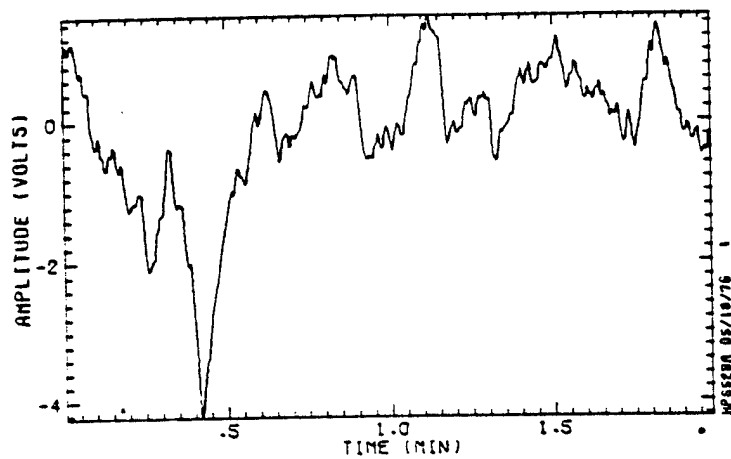
For this particular analysis, ten sections of 1024 points overlapped by 512 points were used. Under these conditions it can be shown, Reference 7, that the variance of the spectral estimate is given by

$$\text{Var} \left\{ S_{xx}(w_k) \right\} = \frac{|S_{xx}(w_k)|^2}{9.104} \quad (6)$$

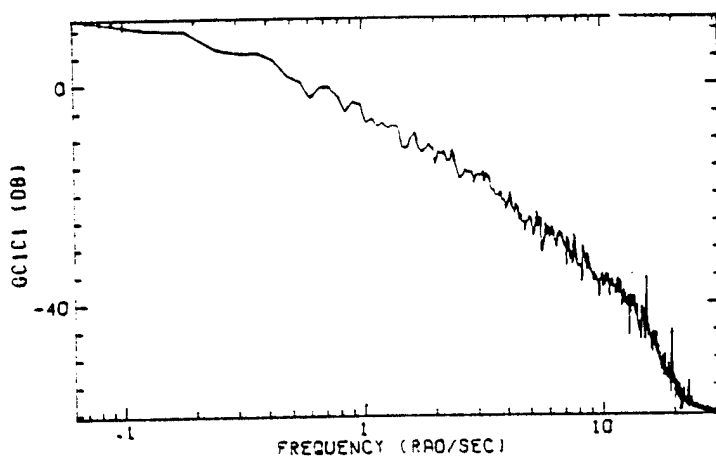
Utilizing this technique to estimate the PSD of the forcing function gives the results presented in Figure 10. The time domain signal $C1$ is shown in the top portion of the figure and the estimate of its PSD ($GCLC1$) is shown in the lower portion of the figure.

In order to investigate the cross coupling in the human operator for the two dimensional tracking task, the closed loop cross-coupled model shown in Figure 11 is utilized. Ordinary, partial and multiple coherence functions can be calculated as described in Reference 7. For notational convenience the frequency dependence of all terms in the following equations has been dropped. The ordinary coherence between input c_1 and output r_1 is given by

$$\rho_{c_1 r_1}^2 = \frac{|S_{c_1 r_1}|^2}{S_{c_1 c_1} S_{r_1 r_1}} \quad (7)$$



A/P FORCING FUNCTION $C(t)$. $FC = .1\text{HZ}$



PSD OF A/P FORCING FUNCTION $C(t)$. $FC = .1\text{HZ}$

FIGURE 10

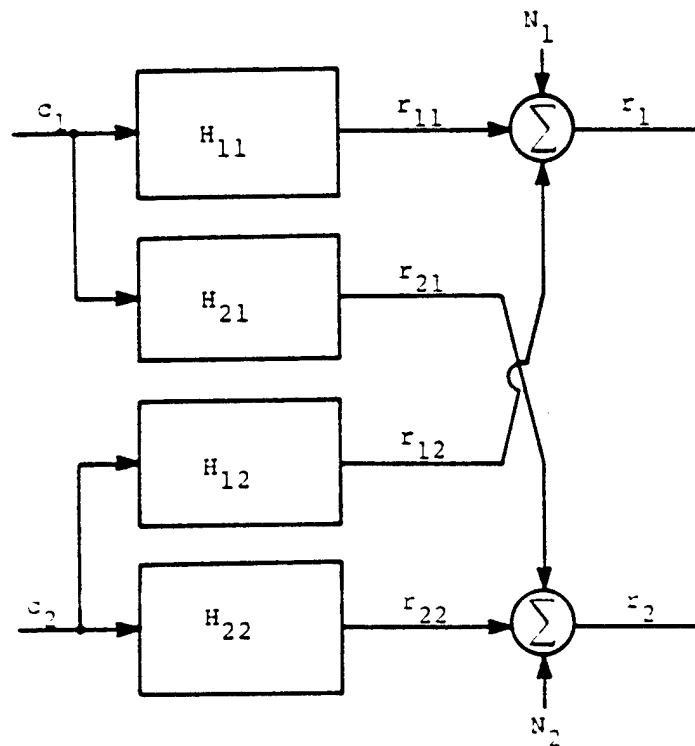


FIGURE 11. Closed Loop Cross-Coupled Model

This function describes the portion of the power contained in the output r_1 which is attributable to c_1 passed through a linear system H_{11} . The ordinary coherence for H_{22} is obtained by changing all the subscripts from one (1) to two (2).

The partial coherence between input c_2 and output r_1 is given by

$$\rho_{c_2 r_1 \cdot c_1}^2 = \frac{\left| s_{c_1 c_1} s_{c_2 r_1} - s_{c_1 c_2}^* s_{c_1 r_1} \right|^2}{s_{c_1 c_1}^2 s_{c_2 c_2} s_{r_1 r_1} \left| 1 - \rho_{c_1 c_2}^2 \right| \left| 1 - \rho_{c_1 r_1}^2 \right|} \quad (8)$$

This function describes the portion of the power contained in the output r_1 which is attributable to c_2 passed through a linear system H_{12} , with all effects of c_1 removed.

The multiple coherence function

$$\rho_{r_1 \cdot c}^2 = 1 - \left| 1 - \rho_{c_1 r_1}^2 \right| \left| 1 - \rho_{c_2 r_1 \cdot c_1}^2 \right| \quad (9)$$

describes the portion of the output power attributable to both inputs via linear systems. For equations 8 and 9, the coherence functions for the other input can be determined, as it was for equation 7, by interchanging the numerical subscripts.

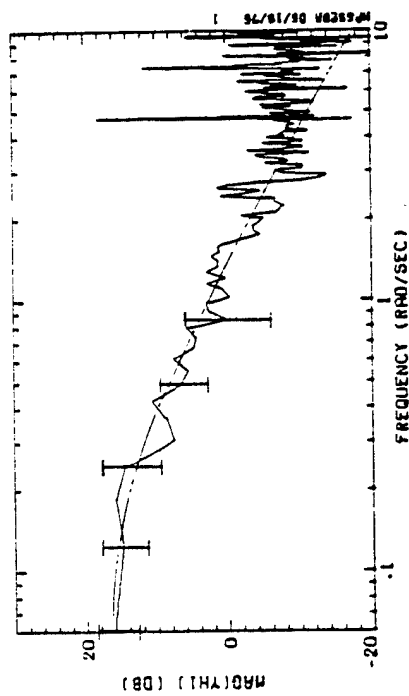
The remnant, that is the remaining output power which is not explained by the linear system, may be found from

$$G_{NN} = (1 - \rho^2) G_{rr} \quad (10)$$

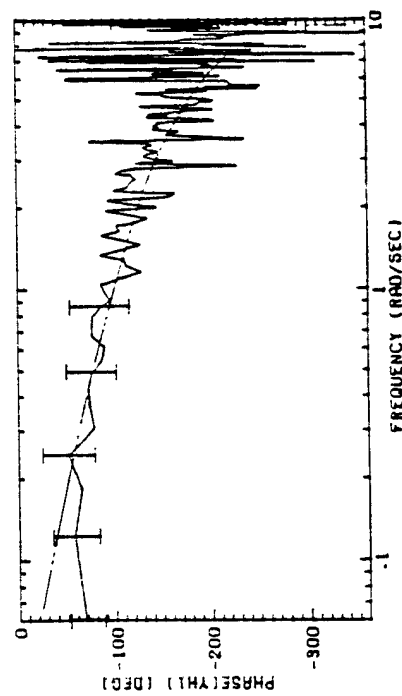
where G_{NN} is the remnant power. When the coherence (ρ^2) is large, that is as it approaches 1, the remnant power can be seen to approach 0.

RESULTS

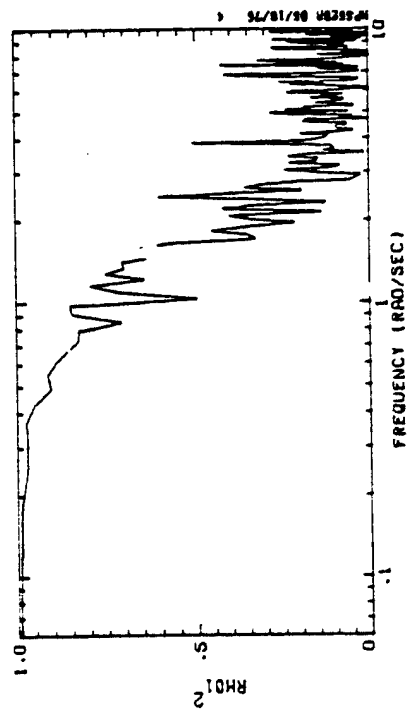
The results of a typical anterior/posterior tracking experiment are shown in Figure 12. The magnitude of the describing function YH1 is given in the top left of the figure. The phase of YH1 is given in the lower left of the figure and the coherence function for this task is presented at the right. The vertical bars indicate the 95 percent confidence intervals for the estimates of magnitude and phase. The phantom curve is a visual fit to the estimated magnitude and phase. The equation of this model is



349



SINGLE CHANNEL TRACKING TASK , R/P PLANE
YH1 DESCRIBING FUNCTION , SUBJECT PV



SINGLE CH. R/P COHERENCE FUNCTION , SUBJECT PV

FIGURE 12. Typical Results of One-Dimensional Analysis

$$Y_{H1}(s) = \frac{6.31e^{-0.29s}}{5s + 1} \quad (11)$$

The coherence function indicates that a linear model describes most of the output power to a frequency just beyond 1 radian per second for this particular task. The main difference between this model and those obtained from visual tracking tasks is the delay time. For tactile tracking, delay times of 250 to 300 msec have been found where delay times of 100 to 150 msec are reported for visual tracking.

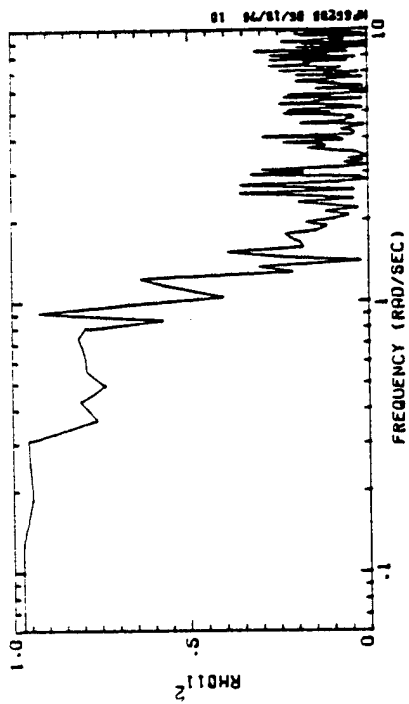
The results of the coherence function analysis for the two-dimensional task are shown in Figures 13 and 14. The results for the Anterior/Posterior (A/P) output (Figure 13) indicate that the ordinary and multiple coherence functions are almost identical. This is due to the extremely low coherence of the A/P output with the Lateral input as shown in the partial coherence function on the right side of the figure. This lack of coherence leads to the conclusion that there is no linear system which can explain the A/P output power from the Lateral input. Figure 14 presents the same results for the Lateral output. Again the contribution of the cross-coupling term in explaining the Lateral output is not significant. This analysis shows that, for the task considered here, the subject is capable of uncoupling the two-dimensional task into two one-dimensional tasks. Therefore, it suffices to model the subject in a two-dimensional tracking task by using two one-dimensional models as shown in Figure 9, represented by an equation of the form given in equation 11.

The parameters of equation 11 will vary as a function of task dimensionality. This can be seen by comparing the open loop describing function magnitude for one- and two-dimensional tactile tracking (see Figure 15). The describing function bandwidth is reduced when the complexity of the task is increased. In addition to that change, the time delay associated with two-dimensional tactile tracking is approximately twice that of one-dimensional tactile tracking, or 500 msec.

Tracking tests accomplished with an L3 paraplegic indicate that his ability to track the signal is as good as that of normals.

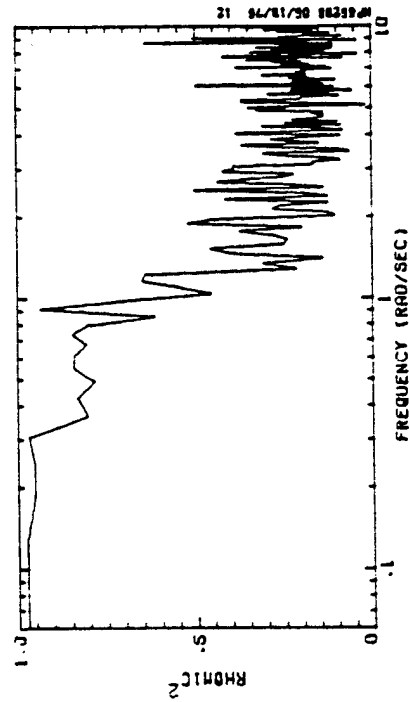
CONCLUSIONS

The feasibility of tracking with a tactile display using electrocutaneous stimulation at the waist has been demonstrated in one- and two-dimensions. Quasilinear models have been fitted to the data. The cross-coupling terms in the model are negligible; hence, for the range of frequencies used as forcing functions in this study, the subjects are capable of decoupling a two-axis tracking task into two independent single axis tasks. Paraplegics seem to be able to track with this type of display as efficiently as normal individuals.

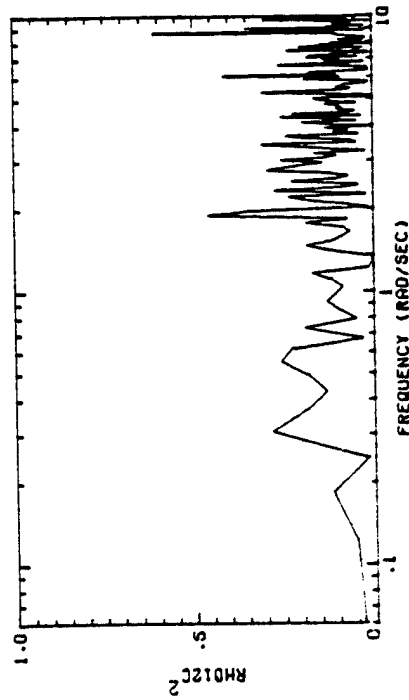


DUAL CH. AP ORD. COHERENCE FUNCTION , SUBJECT PV

351

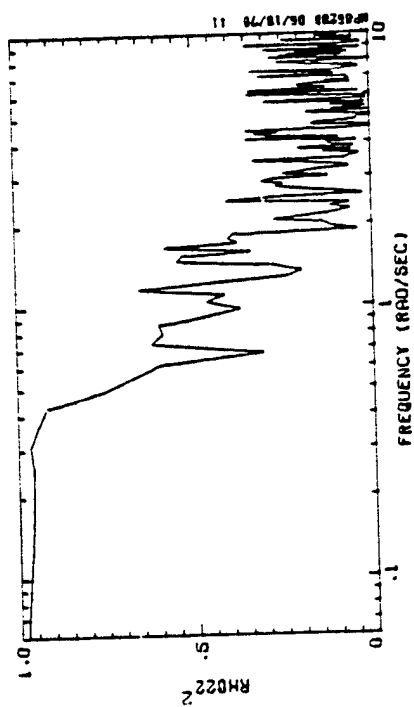


DUAL CH. MULT. COHERENCE FUNCTION , SUBJECT PV



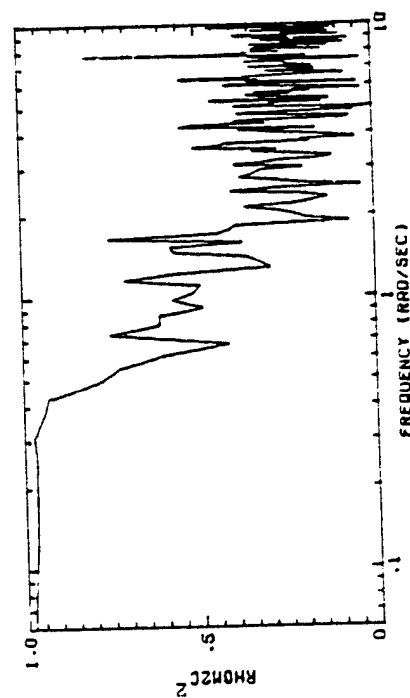
DUAL CH. AP/L PART. COHERENCE FUNCTION , SUBJECT PV

FIGURE 13. Results of Anterior/Posterior
Coherence Function Analysis

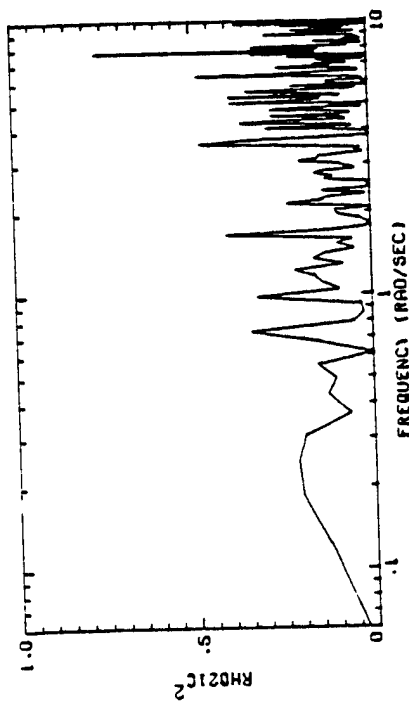


DUAL CH. LAT. ORD. COHERENCE FUNCTION, SUBJECT PV

352

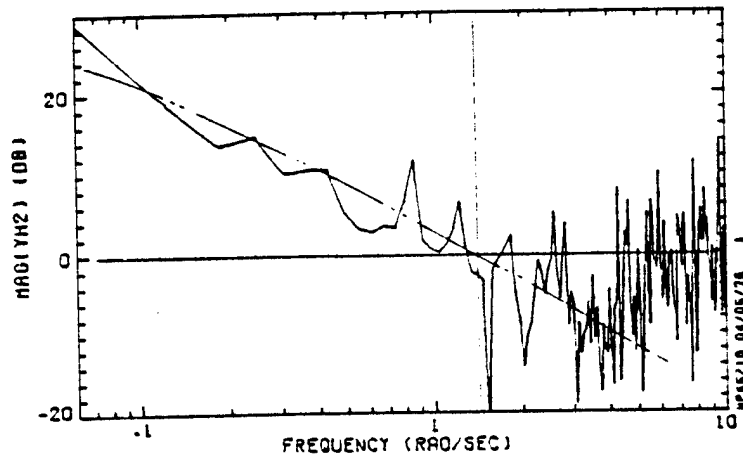


DUAL CH. MIN. COHERENCE FUNCTION, SUBJECT PV

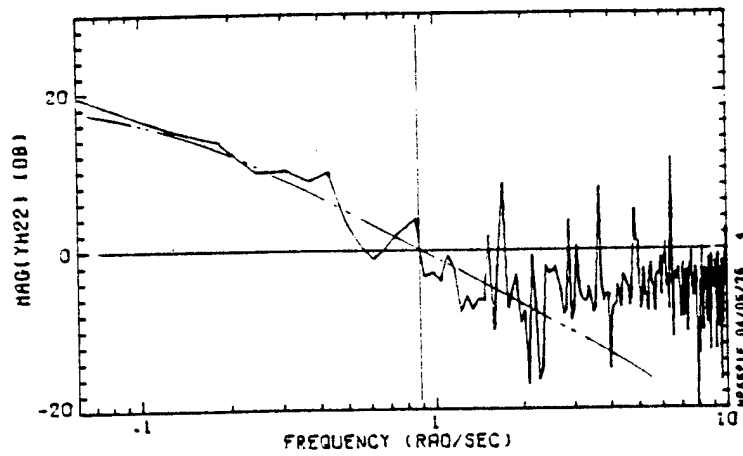


DUAL CH. L/RP PART. COHERENCE FUNCTION, SUBJECT PV

FIGURE 14. Results of Lateral Coherence Function Analysis



SINGLE CHANNEL TRACKING TASK . LAT PLANE
YH2 DESCRIBING FUNCTION . SUBJECT OT



DUAL CHANNEL TRACKING TASK . LATERAL PLANE
YH22 DESCRIBING FUNCTION . SUBJECT OT

FIGURE 15

REFERENCES

1. Hill, J. W., "A describing Function Analysis of Tracking Performance Using Two Tactile Displays", IEEE Trans. on Man-Machine Systems, Vol. MMS-11, No. 1, pp. 92-101, March 1970.
2. Ross, D., Sauneman, R., Levinson, W., Tanner, R. and Triggs, T., "Tactile Display for Aircraft Control", Final Technical Report, Contract N00014-73-C-0031, August 1973.
3. Bensel, C. K., Dzendolet, E., "Power Spectral Density Analysis of the Standing Sway of Males", Perception and Psychophysics, Vol. 4, pp. 285-288, 1968.
4. Saunders, F. A., "Electrocutaneous Displays" in Geldard, F. A. (Ed), Conference on Vibrotactile Communication, Austin, Texas: The Psychonomic Society, 1974.
5. Hampton, R. L. T., "A Hybrid Analog-Digital Pseudo-Random Noise Generator", Proc. Spring JCC Vol. 25, pp. 287-301, 1964.
6. McRuer, D. T. and Krendel, E. S., "The Human Operator as a Servo System Element", Journal of the Franklin Institute, Vol. 267, No. 5, pp. 1-49, 1959.
7. Welch, P. D., "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Trans. on Audio and Electroacoustics, Vol. AU-15, No. 2, pp. 70-73, June 1967.
8. Bendat, J. S. and Piersol, A. G., Random Data: Analysis and Measurement Procedures, Wiley-Interscience, 1971.

A SIMULATOR STUDY OF HELICOPTER PILOT WORKLOAD

REDUCTION USING A TACTILE DISPLAY

An informal report presented to the
Twelfth Annual Conference on Manual Control

By Richard S. Dunn, Richard D. Gilson, Pershing Sun

Mission performance in current tactical Army helicopter operations is significantly limited by high visual workload. Aircrew tasks in low level or terrain flight demand continuous outside visual reference to fly, navigate, and perform mission functions. In nap-of-the-earth or contour flight, the airspace is a ground-defined environment in which operators strive for masking cover, speed and accurate navigation. This kind of visual contact flight leaves no time for scanning display devices in the cockpit. These considerations have generated interest in a simple tactile display concept which may form the basis of an effective information display strategy for Army helicopter cockpits.

This is a brief report of a limited simulation experiment intended to test the feasibility of a tactile display. It had two objectives: to determine whether the adaptive measurement procedure developed by Dr. Pershing Sun at the Avionics Laboratory (Reference 1) could be used to evaluate novel display devices in tactical operations; and to determine whether a tactile display could provide useful reductions in visual workload in tactical helicopter operations.

The laboratory prototype single channel tactile display used for these tests has been described in previous reports to this conference (Reference 2) and is shown in figure 1.

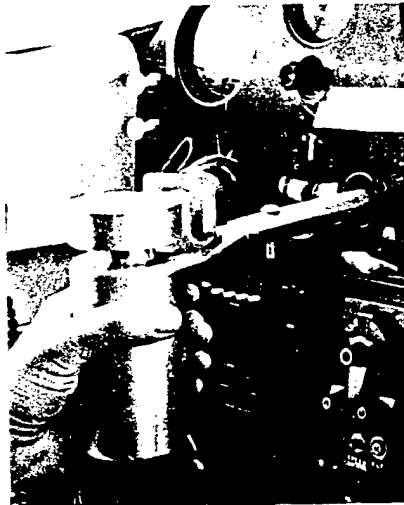


Figure 1. Tactile Display Prototype

The piston moves in both directions to cue a 2-way tracking control movement. It is a compensatory display in use; the operator tracks the output to obtain a centered or flush piston. Tests were conducted at the US Army Electronics Command's Avionics Laboratory, Fort Monmouth, New Jersey using the fixed base helicopter simulator, the Tactical Avionics System Simulator (TASS). The device was installed on the collective lever to provide altitude or power commands as shown in figure 2.

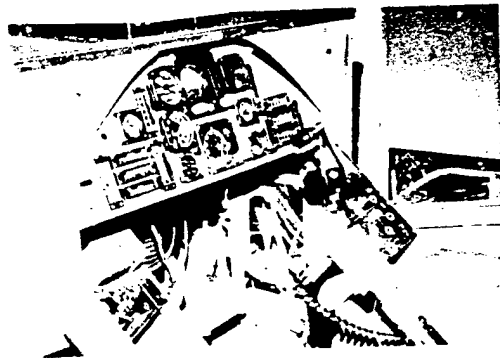


Figure 2. TASS Cockpit with Tactile Display on Collective Control

Two separate tests were performed. The first task was a steep ILS approach with a deceleration phase in the last 400 feet of altitude. The math model for this task represented a UH-1 helicopter. Here, tactile cueing was compared with the normal glide slope command in a four cue electromechanical flight director. A unique adaptive scoring procedure was employed as the dependent variable. Dr. Sun has described this procedure in detail at previous meetings. Briefly, the handling qualities of the vehicle are slowly degraded by the introduction of a pure time delay in the cyclic roll control. The amount of time delay is increased adaptively until a primary task performance measurement circuit indicates that the equally weighted sum of errors in vertical, lateral, speed, and heading (the four commands of the flight director) are at a criterion value. This procedure has the disadvantage of altering the vehicle handling qualities, but the advantage that every subject gives the same level of system performance. He flies just at the acceptable error criterion. In this adaptive measurement, primary task performance alters primary task difficulty until the criterion is reached. The amount of time delay present after reaching criterion is taken as an indicator of workload for the combined display-control-task combination. After practice, five qualified helicopter pilots each made five ILS approaches with the standard visual cue on the flight director and five approaches with the tactile display. Typical results are shown in figure 3.

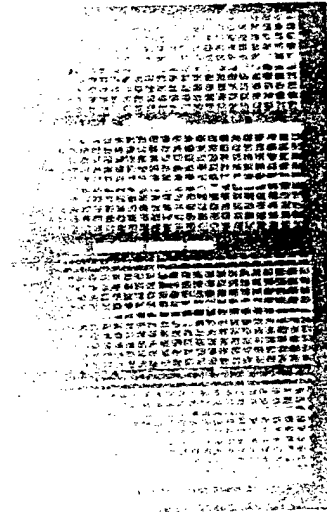


Figure 3. Tracking Command and Time Delay Results

The two tracings on top show the glide slope error (or command) signal: first visual tracking, and below it tactile tracking. Time delay achieved is taken after the build up period, the first 90 seconds, has elapsed. All subjects on all trials showed a greater time delay when using the tactile display. Over all subjects and trials, the increase in time delay using the tactile display averaged 36%. This along with the higher frequency of response to altitude seen in the tracings for the collective command support the hypothesis that visual workload relief is present and is effective at enhancing system performance.

But visual scanning in a flight director is not clearly an example of the kind of visual workload present in tactical helicopter operations. To get closer to the operational environment of interest a second, basically different, simulation was made.

For this task a visual scene was employed, generated by a 300 to 1 scale moving-belt terrain model. The display was a small black and white TV without collimated viewing. The task required hovering over a point at an assigned heading. Changes in altitude were commanded to simulate a tactical bob-up maneuver. The helicopter math model employed represented an OH-6 scout helicopter.

Command information shown in Figure 4 was superimposed over the video display. It was the Integrated Trajectory Error Display operating in the hover display mode (Reference 3). It provides both attitude and position cues which can be employed for hover position tasks. The cross

is a movable indicator of desired ground position. The circle is a fixed ownship position indicator. The line displays velocity, both magnitude and direction, while the small bug is an acceleration cue. For these tests, the left hand vertical band displayed absolute altitude with a collective command on the left inner scale. The heading command was provided by the indicator at the center bottom of the display.

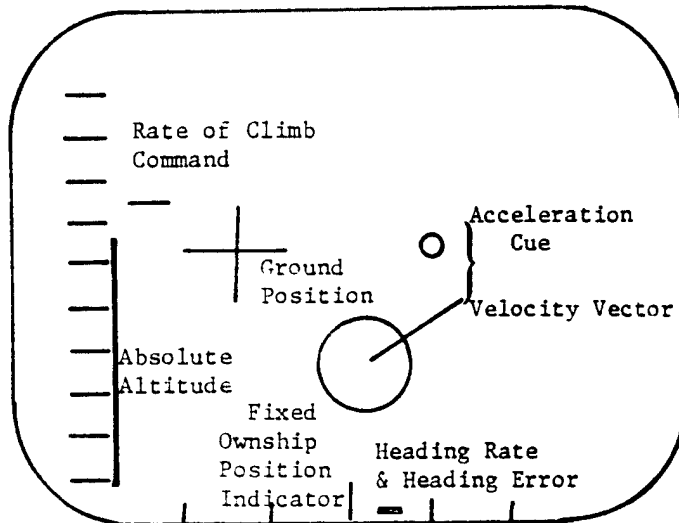


Figure 4. Integrated Trajectory Error Display - Command Information

The task required hovering in position while holding an assigned heading, and altitude. After one minute a maximum performance climb was required to a new altitude with subsequent holding at the new level. This was followed by descent to the original altitude. The change was from 25 feet to 125 feet and back. Auditory warning signals were presented to alert the pilot to upcoming altitude changes.

To estimate the effect of the tactile display on visual workload a visual secondary task was employed as the dependent variable. As before, the adaptive circuits computed primary task performance and provided a weighted sum of error scores. But the output was not used to alter the vehicle handling qualities and no time delays were applied. Instead, the adaptive measurement procedure was used to alter the difficulty of a secondary task by turning the task on and off.

The task involved cancelling lights and was similar to a secondary visual task employed by Kelley and Wargo (Reference 4). Two lamps were mounted in the peripheral field of view so that it was clear when one came on, but a shift in gaze away from the main display was required to determine which one. Lights were illuminated randomly one half second after the previous response. Cancelling required operation of the appropriate left or right rocker switch on the cyclic grip. These were stan-

dard Army intercom or radio push-to-talk switches. When the primary task scoring circuits showed performance within the error criteria, the secondary task was activated. When performance exceeded the criteria, the secondary task was deleted. Thus, the situation represented cross-adaptive control of secondary task difficulty by measurement of primary task performance. As might be expected, the basic flight task proved to be very difficult with the narrow field of view visual display and no motion cues.

Only three subjects obtained sufficient practice in the time available to perform the task reliably without loss of control. It is important to note, however, that the cross-adaptive procedure resulted in stable system performance at or near the error criteria for all these subjects. Each one performed the task six times with the visual cue and six times with the tactile display.

Here the results also favored the tactile display. Overall, the trials with the tactile cue resulted in activation of the secondary task about the same amount of time or slightly longer, but with more than 15 percent more correct light cancellation responses when using the tactile display. This supports the hypothesis of equal system performance with an increase in visual workload reserve.

Unfortunately, scheduling for both the experimenters and the TASS facility prevented testing of enough subjects to solidify these findings. They must therefore be classed as preliminary and only suggestive. But they are encouraging and support the contention that the tactile display concept can produce useful reductions in helicopter pilot visual workload -- useful in the sense that the workload reductions can free visual attention for other productive tasks.

This research was jointly sponsored by the US Army Air Mobility Research and Development Laboratory and the Avionics Laboratory of US Army Electronics Command; and aided by participation of personnel from the Research Foundation of Ohio State University.

REFERENCES

1. Dukes, T.A.; and Sun P.: A Performance Measure for Manual Control Systems. Proceedings of the Seventh Annual Conference on Manual Control, June 1971, 257-263.
2. Gilson, R.; and Fenton, R.E.: A Tactual Pilot Aid for the Approach-and-Landing Task. Proceedings of the Eighth Annual Conference on Manual Control, May 1972.
3. Dukes, T.A.: An Integrated Display for Trajectory Control of Helicopters and VTOL Aircraft. Proceedings of the Sixth Annual Conference on Manual Control, April 1970, 133-145.

4. Kelley, C.R.; and Wargo, M.J.: Cross-Adaptive Operator Loading Tasks.
Human Factors, 1967, 9(5), 395-404.